Gender Differences in Undergraduate Students' Performance, Perception and Participation in Physics

Robyn Claire Annabel Donnelly

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
to the
University of Edinburgh
October 2014

Abstract

Research has been undertaken to obtain a thorough understanding of the existence and degree of gender disparity in students' participation and performance in introductory university physics courses at the University of Edinburgh. The research on this topic has focused on three main subject areas: the proportion of male and female students enrolled in undergraduate physics courses and their reasons for choosing to study this subject, gender differences in student performance and, finally, how students' attitudes and beliefs towards studying physics change after a period of instruction. Gaining an insight into students' attitudes towards studying and learning physics, as well as their conceptual understanding of the topics being assessed, can draw attention to potential areas of weakness which can be targeted in future teaching.

This thesis comprises a comprehensive review of the current situation surrounding male and female participation in the undergraduate physics degree programme at the University of Edinburgh in comparison to other STEM subjects, as well as a description of factors potentially influencing the gender performance in physics. With respect to student performance, conceptual understanding tests have been used as evaluation tools to measure the effectiveness of introducing interactive engagement, such as Peer Instruction, into teaching environments in order to improve student performance, as well as a means by which male and female learning gains could be compared. Results indicate that female students show a lower level of conceptual understanding of Newtonian Mechanics than male students when entering the degree programme, and that this gender difference remains after a period of instruction. Qualitative interviews highlight the preconceptions of first year undergraduate physics students with regards to Newtonian concepts of force and motion and demonstrate the range of misconceptions held by both male and female students.

The research presented here compares male and female performance on different forms of assessment; coursework, laboratory assessments, examinations and peer instruction in-lecture questions. Results indicate that while examination scores show no distinct gender trends, female students show consistently higher coursework scores compared to males across physics, chemistry and biology first year courses. Analysis of Peer Instruction questions implemented in the introductory physics lectures suggest that such teaching methodologies have had an overall positive effect on class performance, although there is evidence that differences exist between male and female performance on individual questions.

Students' attitudes towards learning physics have been measured at undergraduate level in order to evaluate the level of 'expert-like' thinking of first year undergraduate students. One notable finding of this study has been the lack of decline in the 'expert-like' thinking after a semester of teaching in recent years, where previously a decline had been witnessed in this expert-like thinking. This result coincides with a change in the format of lectures to a 'flipped-classroom' approach and may have implications for the introduction of new teaching methods. As well as focusing on the progression of undergraduate students' attitudes, this study has evaluated UK academics' attitudes towards physics. This has enabled a UK level of 'expert-like' thinking to be established, with gender differences between male and female academics identified. Students' opinions of the transferable skills gained and their experiences during their degree programme are discussed. Each of the gender topics discussed in this thesis has provided a deeper insight into gender differences in student attainment at undergraduate level which could have implications for the further improvement of future courses.

Lay Summary

Females are highly under-represented in university physics courses. The deciding factors that contribute to the decision to pursue physics are discussed using data collected from students at the University of Edinburgh. For those students who have made the decision to enrol in the physics degree programme there exists evidence of a gender gap in performance.

One of the main aims of an undergraduate physics degree programme is to improve students' overall understanding of physics concepts and develop their ability to transfer this knowledge to solving problems presented in a diverse range of contexts. The most common way of tracking the progression of students' attainment is through the comparison of course grades. Such analysis has reported the existence of gender differences in physics students' performance on different courses, a result which opens up many avenues for further investigation. This can ultimately influence teaching methods used to target problem areas, particularly in the first year of university, which can benefit both genders.

While the use of interactive teaching methods has shown an overall improvement in students' understanding after a semester of teaching at the University of Edinburgh, differences exist between the performance of males and females. Comparisons of male and female learning gains are possible through the use of concept inventory tests which measure the level of understanding of a particular physics concept using multiple choice questions. These questions are created specifically to target common misconceptions held by students. One of the foremost examples of gender discrepancy is in relation to concepts of force and motion, where it was found that females had consistently lower scores than males. These responses were further explored with respect to the specific misconceptions of students. Gender differences in students' responses to a question and their

confidence in their answer were two emerging themes that emerged from listening to students' problem solving strategies during interviews. For example, it was seen that females were much more likely to show a high degree of confidence in their incorrect answers than males.

Student performance on university courses was evaluated using a variety of different assessment methods. This thesis compares students' scores on coursework, examinations, laboratory reports and in-lecture questions with respect to gender. An evaluation of these results found evidence that females invariably outperformed males on continual assessments, both in physics and other first year science courses, but no consistent gender trends existed for examination scores. In-lecture questions, which students answered using personal response clickers, showed high learning gains by both cohorts, although performance levels and gender gaps differed depending on individual questions.

The final area of research in this thesis explored students' perceptions of their degree experiences and their attitudes towards learning and studying physics. Overall, students showed relatively high levels of thinking towards the subject compared to experts in the field. The change in students' attitudes towards study over their first year of university was measured and indicated that, while students previously showed a decrease in 'expert-like' thinking after two semesters, recent years have showed no such drop in students' positive attitudes towards studying physics. Discussions with students, both in the early years of their degree and on graduation from the degree programme, offered interesting insights into areas of the courses which they found successful or alternatively in need of improvement. Such results have implications for planning courses for future students.

Declaration

Except where otherwise stated, the research undertaken in this thesis was the unaided work of the author. Where the work was done in collaboration with others, a significant contribution was made by the author. The candidate confirms the appropriate credit has been given within the thesis where reference has been made to the work of others.

Parts of this work, discussed in Chapter 3, have been written up for publication in the European Journal of Physics:

S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch and N. R. Walet. Gender differences in conceptual understanding of Newtonian mechanics: A UK cross-institution comparison. *European Journal of Physics*, 34(2):421-434, March 2013.

Robyn Claire Annabel Donnelly
October 2014

Acknowledgements

First and foremost I would like to thank my supervisors Professor Cait MacPhee, Professor Judy Hardy and Professor Simon Bates for their invaluable advice and guidance throughout my studies. I am also grateful to all the past and present members of the Physics Education Research group at the University of Edinburgh, Alison Kay, Marsali Wallace, Katherine Davies and Dr. Ross Galloway for their insight and revelations into Physics Education Research and for making this an enjoyable experience. Without them this PhD study would not be what it is. I also thank collaborators from other institutions for their contribution to the research included in this thesis.

A special thank you to my family and friends who have listened patiently to updates about my research and have offered their endless support and feedback which has meant so much to me. My parents Pauline and Patrick Donnelly and my sister Candice, without whose love and encouragement I would not be where I am today.

A final thank you to the University of Edinburgh and the Principal's Teaching Award Scheme for funding and providing me with the opportunity to undertake this research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Motivation

The under-representation of women in physics is of considerable concern at all stages of the academic pipeline. Despite high achievement in school qualifications, the number of girls making the decision to pursue physics in tertiary education is low compared to many other sciences. Multiple factors can contribute to a student's decision to study a subject either at school or university [1, 2]. In order to encourage more students to study physics at a tertiary level, first we must investigate the motivation for students' interest in the subject, the reasons which may cause that interest to wane and why this ultimately leads to a disproportionate number of males entering physics. Gaining further insight into students' reasons for their degree choices may help inform instructors of how STEM disciplines can be made more attractive to female students. In addition to the gender disparity in participation levels, evidence suggests that, females who have made the conscious decision to enrol on a university physics course underperform compared to males in physics [3, 4]. Exploring the existence of this performance gap and identifying areas which could be targeted to minimise the gender difference in attainment are important for improving gender equality in physics education.

Why there should exist a gender difference provokes ongoing speculation. Some studies have suggested that this may be a consequence of innate cognitive and psychological differences between males and females, for example gender differences in spatial or visual reasoning [5]. There exist conflicting results, with many arguing that small biological differences are not sufficient to explain the under-representation of women or observed performance differences between genders [6, 7]. The focus of the research presented in this thesis is on identifying where gender differences exist in university education, as measured by student performance and attitudinal levels.

This thesis examines three key features in the university experience in order to gain a wider picture of gender differences in the undergraduate physics population and an understanding of how multiple factors contribute to the overall gender disparity, specifically: participation, performance or attainment and students' perceptions and attitudes to learning. Each of these areas has been investigated to provide further insight into the gender issues surrounding physics education at university. In this thesis, the focus is placed on exploring measurable differences in male and female students' conceptual understanding in physics, the change in the magnitude of these differences as students progress through their degree, and on investigating their experiences during their undergraduate studies. Results from such research then can be used to inform future teaching methods and strategies to support students' academic studies.

While Physics Education Research (PER) is a developing area in the UK, many North American institutions have well established research groups. The majority of studies examined in the literature have focused on gender differences in secondary school education or at North American universities, which have different educational and instructional programmes to those in the UK. North American universities also often have considerably different proportions of male and female, and major and non-major, students taking first year courses. These marked differences in the composition of introductory physics course cohorts between North American and UK universities offer us the opportunity to explore similarities and differences in student participation and attainment between the two.

There is extensive literature on gender issues in science education, in particular the discrepancies at secondary school level [8, 9, 10]. This study does not seek to measure potential effects occurring in primary and secondary education that may have influenced the already existing participation gap, but acknowledges that there may exist differences in the backgrounds and prior exposure to physics among the incoming cohort of students as they transition into university. Therefore, the research questions focus on the gender discrepancies seen in our undergraduate courses. Further understanding of participation gender gaps can be gained by asking if there is an evident gender difference in reasons for students' interest in the subject or their expectations of their learning environment. The issue of undergraduate performance consists of several different areas including students' conceptual understanding, misconceptions and the effect of assessment type. Whether male and female students exhibit similar growth in conceptual understanding after a period of instruction, and whether the difference between gender performance is dependent on the format of the assessment administered, also is approached in this study. Finally, this research questions how the attitudes of undergraduates towards studying and learning physics change over a period of time.

Each of the chapters in this thesis will provide information on the motivation behind and results from each area of the study along with comparisons with previously published data. In this introductory chapter, relevant literature will be discussed with specific reference to the overarching research aims of this thesis.

1.1 Student participation in STEM subjects

It is widely recognised that there is a need to encourage more students to pursue further study in all STEM (Science Technology Engineering Mathematics) disciplines. The decline in levels of engagement in STEM subjects has led to the initiation of local and national projects to raise the profile of STEM careers across the UK [11, 12].

Ongoing concerns are emerging regarding a gradual downturn in students' interest in studying physics in particular. The number of students taking physics at secondary level in the UK has decreased dramatically over the last few decades. In 1985 the number of entries to A-level physics stood at 47,000 (approximately 7% of all A-level entrants), falling to less than 28,000 (approximately 3.5% of all A-level entrants) by 2006 [13]. As well as a reduction in the total number of students studying physics at secondary level, there is strong evidence to indicate an under-representation of girls in physics courses throughout Britain. A report published by the Institute of Physics (IOP) about “Girls in physics” showed that while the total number of pupils completing A-level physics increased by 5% between 2011 and 2012, the proportion of girls remained relatively unchanged, going from 20.5% to 20.9% [14]. Additionally, the proportion of females taking physics A-level examinations has remained much lower than that in other STEM disciplines such as mathematics (41%) [15], chemistry (47%) and biology (57%) [14]. Retention is also a factor in the gender imbalance, particularly the retention of female students. For example, in 2010 the proportion of females dropped from 24% to 21.5% from AS level to A-level [15] and in Scotland from 27% in Higher physics [16] to 23% in Advanced Higher [17].

There have been reports that enrolment in mixed or single sex schools could have an effect on the gender uptake of physics courses by school pupils and on females' confidence [18]. Gill and Bell noted that males studying at a mixed school and females in single sex education were more likely to choose to pursue physics at A-level [1]. The authors commented that this may be due to both teachers' and students' expectations of it being a more male orientated subject. A recent study by the IOP investigating gender participation in physics at secondary school similarly found that girls attending a single sex school were 2.5 times more likely to study physics at A-level [19]. Alarmingly, in 2011, 46% of secondary schools in England had no girls taking A-level physics [19]. This IOP report, along with other studies looking at the factors affecting participation and attitudes to studying science and mathematics, comment that teachers and family members play a crucial role in encouraging and supporting girls to study science [19, 20].

The gender participation problem is not confined to secondary education and it is on the situation at university level that this thesis is focused. The number of accepted applicants to degree courses in physics has been rising in the last few years, but still remains relatively low, with 4,000 applicants in 2011 (0.93% of accepted applicants to all degree courses in 2011) [21]. There exists a noticeable gender gap in the number of students enrolling in undergraduate physics degree programmes in the UK. Statistics released by the IOP indicated that the proportion of females entering first year undergraduate physics has remained between 18-20% for the last 15 years [22]. Retention rates at university are also an area of unease. The decline in physics students across secondary and tertiary levels has been described as a '*leaky pipeline*' [6]. Students, particularly women, who were initially interested in pursuing a career in physics and other sciences, are lost at various stages along this pipeline: in the transition from secondary school to university; over the course of their university degree; and after completion of a science degree in entering a non-scientific career. In the

UK many steps have recently been taken to encourage girls to pursue degrees in the sciences including the introduction of projects such as Women into Science and Engineering (WISE) [23] and Girls into Physics [24] which aim to promote science in formative years.

Blickenstaff refers to several possible explanations for the gender discrepancy in the uptake of STEM subjects and careers [6]. The influence of social and cultural stereotypes, teacher and family encouragement, the absence of female role models in the sciences, and the teaching environment of the science curriculum which some suggest favours male students, have all been alluded to as potential contributors to the observed gender gap in participation [4, 6]. An increasing volume of research is being undertaken to determine what factors influence attitudes and interests in science, the selection of physics as a degree course, and factors affecting retention rates, both in the transition between secondary and tertiary education and during university [25]. Whether there is an identifiable gender dimension to these factors is an emerging issue and is touched upon in Chapter 7, in which students' reasons for choosing to take the first year undergraduate physics course at the University of Edinburgh are explored to determine if gender differences exist.

1.1.1 Stereotype threat

In addition to tracking student participation at different levels of study, some studies have looked at social-psychological influences and effects of stereotypes on the observed gender discrepancies in participation and performance in STEM subjects [26, 27, 28, 29, 30]. This is an extensive area of study with previous research claiming that gender preconceptions and stereotypes can affect children's attitudes towards science as early in their education as primary school [31]. McAdams believes that stereotypical images formed at a young age establish "*attitudes and social expectations which are seldom modified by subsequent experience*" [31] and these stereotypes can go on to influence whether a child chooses to continue studying a subject. Several studies have been done in the last fifty years about students' views of what it means to be a 'scientist' [32, 33]. When asked to describe or draw a scientist the majority of students visualised a "*white-coated man in a laboratory*" with very few suggesting that the scientist may be female [31, 34]. Despite continued promotion of science as a career, this stereotype of science as a male dominated field still persists.

As well as its potential influence on the uptake of STEM subjects by students, stereotype threat has been linked to performance outcomes. A growing area of research discusses the under-performance of minority groups due to negative stereotypes relating to intellectual ability [29, 35]. In a paper published in 2010, Miyake *et al* looked at how a simple intervention consisting of a short writing exercise could affect the performance of women on a Force and Motion Conceptual Evaluation (FMCE) test [36] as well as an end-of-course examination [37]. It was suggested that women may feel extra pressure to perform highly in case a negative performance in an assessment may affirm performance stereotypes, that men are better than women in the sciences. It was also noted that their perception of their performance may continue to have a profound effect on their performance on further assessments. This study, undertaken with students at the University of Colorado, implemented the intervention at the start of the course and again a week before the midterm assessment. In the 15 minute writing exercise students wrote about the values which were most important to them, such as family, friends and procurement of knowledge, and which were entirely unrelated to the context of the course being taken. The authors stated that when students "*affirm their core values in a threatening environment, people re-establish a perception of personal integrity and worth, which in turn can provide them with the internal resources needed for coping effectively*" [37]. The exercise was introduced to students not as a gender issue but as a way to improve their overall course performance. Students who completed the affirmation exercise showed both a decrease in the gender performance gap on examination grades (and a positive shift in grade distributions) as well as a complete elimination of the gap on the FMCE. While a decrease in the gender gap was witnessed, it had the negative outcome of disadvantaging males and decreasing male exam scores. This values affirmation exercise was administered to students at the University of Edinburgh, the results of which are discussed in Chapter 3.

Similar studies carried out looking at differences in male and female performances on maths diagnostic tests have shown improvements in the gender gap if an affirmation statement is presented prior to the assessment [26, 27]. A study by Martens *el al* investigated the effect of including a self-affirmation exercise on females' mathematics performance on questions dealing with spatial rotation [26].

Women who were introduced to a negative stereotype underperformed compared to males, whilst women who completed a form of self-affirmation performed equally to males. The inclusion of the affirmation exercise did not show any effect of increasing the male performance level. Claims that the gender gap could be completely eliminated through the introduction of such interventions as the values affirmation exercise were not universally witnessed [38]. The concept of 'stereotype threat' was further examined by Kost-Smith *et al* by repeating the values affirmation exercise with students in another semester [39]. The study was carried out in the same course as the original study with the same course instructor. While they saw a reduction in the gender gap in exams which was consistent with the original study, they did not see a replication of the FMCE results from the first study. Females in the control group statistically outscored both females in the affirmation group and males in the control group.

As well as the influence of stereotypes on a person's confidence and potential performance, students and academics may also be affected by '*Imposter Syndrome*'. The term '*Imposter Syndrome*' is used to describe the psychological phenomenon in which someone is unable to internalise personal achievements irrespective of external evidence [40, 41]. Rather than associating their success with intellectual ability or competence, those experiencing the Imposter Syndrome can feel a lack of confidence or guilt, often attributing their success to luck or chance. It has been theorised that high-achieving females are more likely to feel that they are undeserving of their success and adopt a mindset that they are an 'imposter' or 'fraud' in their work. In a study by Clance and Imes the beliefs of 150 women were investigated. They witnessed the prevalence of the feeling of not being intelligent enough for their current role [40]. Interestingly, they commented that "*In our clinical experience, we have found that the phenomenon occurs with much less frequency in men and that when it does occur, it is with much less intensity*", suggesting a gender difference in attitudes or confidence.

Stereotype threat may result from an overarching impression that science is a stereotypically male domain. Increasing the prominence of role models may help to overcome this barrier. A lack of female role models in physics, and science in general, emanates from the fact that males comprise a significant proportion of those employed in academia and science and technology related fields [42]. This low proportion of women results in a possible misconception that science is stereotypically a male domain. Family members, teachers and public figures can all have an impact in encouraging more females into STEM subjects. The presence of high profile women in academia and science professions can have the affect of altering girls' perceptions of possible future careers [43]. However, Blickenstaff concluded that examining "*issues like the presence of role models in science is a way of looking at the environment that girls and women encounter as they learn science. It seems that the presence of role models in a science or engineering department would be unlikely to fix the problem of under-representation of women, but could be one part of a solution*" [6]. He makes it clear that although the lack of female role models may have an effect on girls' choices of future subjects to study at high school, or even choice of future careers, on its own it cannot be held solely responsible for the gender disparity in science participation.

1.2 Performance and attainment in physics

The first year of university is a key period in students' education. During this year students build on their previous school knowledge and develop their conceptual understanding of key physics concepts. Students are exposed to a new learning environment and teaching methods which differ considerably to those previously experienced. This means that students need to adopt new learning strategies and partake in more independent study. It is important to identify students' initial level of comprehension and identify potential misconceptions. Studies have investigated the idea that teaching strategies and pedagogical approach have a measurable effect on students' overall performance [44, 45]. In this thesis, the progression of students' understanding of physics concepts, and Newtonian mechanics in particular, is explored. This is examined both from the perspective of students' understanding of physics concepts upon entry to university, measured using conceptual tests such as the Force Concept Inventory (FCI) [46] or Force and Motion Conceptual Evaluation (FMCE) [36], and from gender gaps in performance in assessed coursework and examinations. In the following section the use of conceptual inventories to measure student performance will be discussed along with a review of literature discussing the effects of teaching and assessment methodologies on student attainment and gender performance differences.

1.2.1 Concept inventories

The development of concept inventories in science subjects has had a significant impact on testing methods undertaken in science education research [47]. The primary purpose of research-based inventory tests is their use by instructors to gauge students' understanding, or the change in this understanding, of fundamental concepts within an area of study. By determining an initial reference point to students' understanding, lecturers and course organisers are able to test the level of effectiveness a specific course has had on students, as well as judge how effective any changes to teaching methods have been.

One of the most extensively employed tests of conceptual understanding is the Force Concept Inventory (FCI) which was developed by Hestenes *et al* to measure students' understanding of Newtonian mechanics [46]. This diagnostic instrument has been used as a benchmark for the creation of a range of concept inventories employed in science education. There exists a diverse range of instruments targeted at different subject areas [48]. In Chapter 3 of this thesis, results from the use of three diagnostic tests, the FCI [46], the FMCE [36] and the Brief Electricity and Magnetism (BEMA) test [49] are discussed. Results from the use of these instruments in previously published studies are discussed in the following sections. The design and validation of each of these instruments will be discussed in more detail in Chapter 2.

1.2.2 Interactive engagement and Peer Instruction

Growing evidence exists to suggest that specific teaching methodologies can increase students' learning more than traditional lecture formats [50, 51]. In a paper by Hake, in which he investigated the effects of different instructional methodologies, he defined interactive engagement methods as "*those designed at least in part to promote conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors*" [44]. For comparison, traditional courses were classified as those making "*little or no use of IE* [interactive engagement] *methods, relying primarily on passive-student lectures, recipe labs, and algorithmic-problem exams*" [44]. Hake used data from 6,542 high school and university students from 62 introductory physics courses to test the correlation between pedagogy and student scores on conceptual tests including the FCI [46] and Mechanics Diagnostic test (MD) [52], as shown in Figure 1.1 [44]. On this figure Hake plotted the percentage gain against percentage pre-test score from the MD or FCI for each of the high-school, college or university courses. Each course was defined as a 'traditional' course or a course that made use of 'interactive engagement'. Drawn on this figure are slope lines indicating regions of low, medium and high gain. He found that the average normalised gain[1] in those that followed interactive engagement techniques (<g>=0.48±0.14) was significantly higher than traditional lecture courses (<g>=0.23±0.04), suggesting the efficacy of interactive teaching methods in improving student learning compared to traditional methods. Although, as seen in this plot, not all courses taught using interactive engagement methods examined in Hake's study demonstrated higher learning gains than traditional courses, this result has been noted in studies undertaken in other disciplines [53, 54].

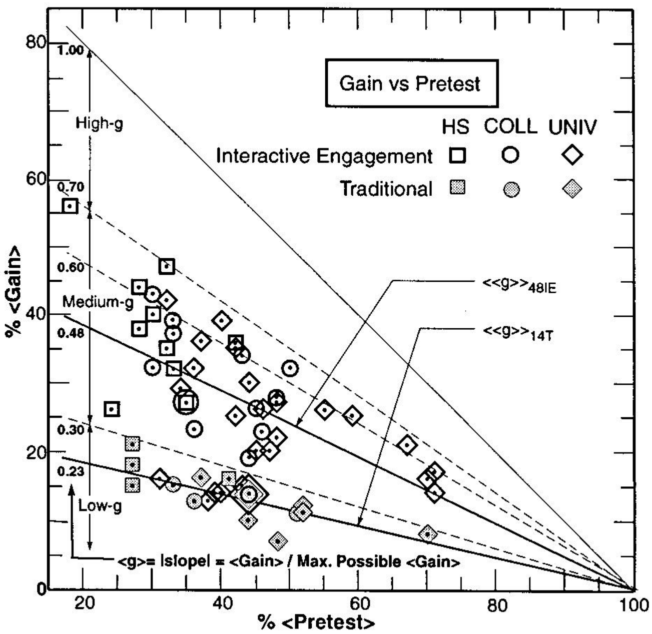


Figure 1.1: Relationship between average gain and average pre-test score for 62 courses with a student population of 6,542 [44].

[1]Normalised gain is defined as the change in score from pre-test to post-test as a fraction of the total possible increase in score.

Research conducted by Lorenzo *et al* at Harvard demonstrated the effectiveness of such interactive engagement in closing the performance gender gap [55]. They employed the FCI to investigate the influence of different types of teaching instruction on the gender gap in introductory physics over a seven year period between 1990 and 1997[2]. Their study focused on integrating Peer Instruction [50] and collaborative problem-solving techniques into the lecture environment of an introductory calculus-based physics course taken by non-major students. They formed a comparison between three separate groups: traditionally taught classes, those that were partially interactive (Peer Instruction in lectures but traditionally taught tutorials) and those classed as fully interactive (Peer Instruction in lectures and group problem solving in tutorials). When students commenced the course there existed a statistically significant gender gap of 9-15%, with males performing more highly than females. After a semester of teaching they found that post-test FCI results indicated a decrease in the gender gap for both partial and fully interactive classes, with the gender gap in fully interactively taught courses no longer statistically significant. Female students also showed a higher absolute gain than their male counterparts. Furthermore, they commented that particular methodologies, such as the interactive engagement style of lecturing using Peer Instruction introduced by Mazur [50], were more effective than others at reducing the gender disparity in performance.

A study undertaken by Pollock *et al* at the University of Colorado showed that the introduction of interactive engagement methodologies in calculus-based physics courses was not sufficient to completely eliminate the gender gap in performance on the FCI [3, 56]. Despite noticeable improvement between pre-test and post-test scores by both cohorts, the gender gap was not fully closed, and, despite a few cases arising where the gender gap was reduced, no statistically significant reductions were seen in the gender gap. In the case of some courses, the gender gap was statistically widened. Comparing this study to that done at Harvard, two differences were noted: The FCI pre-test scores of the students presented in this study were much lower than at Harvard, and the Force and Motion Concept Evaluation (FMCE) test [36] was used in place of the FCI.

[2]The gender performance gap can be defined as the difference between male and female performance and is arbitrarily defined as positive or negative depending on whether it is male score minus female score or vice versa. Throughout this thesis the gender performance gap has been defined to be the average male score minus the average female score.

The authors proposed that differences in background preparation in physics and mathematics of incoming students may partially explain the evident gender disparities in performance and attitudes at university level. They speculated that instructor effects, such as different levels of implementation of techniques may contribute to the observed differences between semesters. A meta-study by Madsen *et al* provides a concise review of studies undertaken to measure the gender performance gap [57]. Results from courses have shown that the implementation of teaching methods have in some cases resulted in the decrease of, or even elimination of, the gender performance gap, although not all such results were replicated in subsequent studies.

A form of formative in-lecture assessment is the use of electronic voting systems (EVS). The introduction of electronic voting systems in science courses has encouraged an increase in peer discussions amongst students both during and outside classes and has been shown to increase student performance in different subject areas [51, 58, 59]. Students are presented with a personal EVS device and during the course of a lecture the instructor presents a question to the class and asks students to individually vote on what they believe to be the correct answer. After this initial vote students are then encouraged to discuss the question and their choice of answer with neighbouring peers before participating in a second round of voting. It has been noted that increases in students' understanding of topics can be improved both through listening to explanations from peers as well as through actively explaining concepts to other students [60]. Presented within Chapter 6 is a review of the use of interactive clickers in first year physics lectures at the University of Edinburgh.

When considering the effectiveness of new teaching methodologies it is important to recognise that different instructors may execute these techniques differently with regard to interactive engagement methods such as Peer Instruction or collaborative problem solving [61]. The potential influence of instructors on the changes in students' performance as measured by conceptual tests is also noted in a study by Docktor and Heller. Docktor and Heller presented a decade worth of results from the use of the FCI in 40 classes with more than 5500 students and 22 different course instructors [62]. Averaged over all courses, they too witnessed a gender gap prior to instruction, with males outperforming females by more than 15%. After a semester of instruction, the gap decreased marginally to 13%, remaining statistically significant. They did however witness a wide range of changes in the gender gap between pre-instruction and post-instruction. The change in the gender gap ranged from -8% (narrowing of gender gap) to +7% (widening of gender gap), and they hypothesised that this was due to the influence of different course instructors in different semesters. Unlike Mazur *et al* [50], Docktor and Heller observed a significant decrease in the existing gender achievement gap only for those students with higher pre-test and post-test scores, with no significant decrease for those with lower pre-test and post-test scores. Despite both studies analysing courses which employed Peer Instruction and interactive engagement techniques, a huge discrepancy was observed in their results. This highlights the potential importance of instructor and student cohort factors.

The introduction of interactive engagement methods in first year physics classes at the University of Edinburgh has resulted in consistent high learning gains over the course of a semester, as measured by conceptual tests [63, 64]. The difference between undergraduate populations in US and the UK offers the opportunity to observe the gender imbalance in students from a different academic background to those reported in previous studies. Results of first year introductory physics students' performance on concepts inventories are discussed in Chapters 3 to 5.

1.2.3 Pedagogy and assessment

One of the key features of the research presented in this thesis is the comparison of different assessment types in the exploration of gender differences. The view that females are able to perform more effectively through group discussion and interaction, while male students favour a teaching technique that promotes a structured learning environment and one that enables them to work more independently, is one that has been extensively researched [65, 66]. Stewart stated that "*physics has traditionally been taught in an abstract rule-dominated way, which appeals more to boys than to girls*" and that girls are more likely to be influenced by the "*perception of subject difficulty*" than males [67]. The author reported that, after sampling 128 A-level physics students, results on GCSE exams proved to be more of an influence on the choice of A-level subjects for females than males. Results suggested that a move towards a more interactive and collaborative learning environment may not only improve students' performance, but may benefit the learning approach of women in particular.

Over the course of their undergraduate degree, a student's performance is assessed through a variety of different methods, including continual assessment and final examinations. The test performances of males and females has shown differences depending on the type of assessment administered [3, 62]. This gender bias in the format of assessment is not unique to physics but has been noted across several disciplines [66, 68, 69]. For example, a study by Elwood noted that females showed higher performance levels than males in coursework at GCSE level across several disciplines including English, Mathematics and Science [66].

Different forms of assessment type at undergraduate level in physics were examined by the University of Colorado in a study that looked at coursework and examination grades separately to see if there was a gender bias [3]. For each of the seven semesters examined, there was no significant difference in the total course grade. Despite this, males were recorded to have outperformed females by an average of 5% on examinations and females outperformed males on coursework by approximately 5% [3]. In an earlier study, the authors note that in some cases coursework assignments are designed to be collaborative with little time dependence, whilst exams are individual and involve an element of competition with a specific time constraint [56]. At Edinburgh, students are encouraged to work on problems within peer groups during course tutorials prior to submitting individual coursework assessments. Docktor and Heller also examined potential differences between male and female performance on overall course grades [62]. Males on average scored 1.5% higher than females in undergraduate assessments, but this gap increased slightly to 3.9% if only examination marks were considered. They suggested that this increase may be due to the fact that the final course mark takes into account lab reports and participation, and therefore may be influenced by student diligence. A similar comparison between gender performances on different types of assessment was examined using data collected at the University of Edinburgh and will be discussed later in Chapter 6 of this thesis.

Steinberg and Sabella explored the difference between students' answers on a multiple-choice diagnostic test and exam problems [70]. While there was some correlation between exam performance and performance on FCI questions relating to the same context, there were questions on which student performance did not match. The authors hypothesised that one of the reasons for this may be that the FCI relies on multiple-choice answering, whereas the end-of-course exam comprises open-ended questions. They stated that, while most of the incorrect exam responses related to possible answer options on the FCI, in some cases students answered correctly on the exam but provided incorrect reasoning. They commented that "*it would not be surprising if the reason that some of the students answered differently on the FCI and the exam was that the FCI triggered responses (right or wrong) that would not have been produced by the students on their own*" [70]. This idea of students' answers to multiple-choice questions not reflecting their true understanding or reasoning of a topic is particularly significant. When interpreting class scores instructors need to appreciate the limitations of such diagnostic tests in quantifying students' understanding, as well as the need for further investigation into qualifying individual misconceptions about certain concepts. Literature has also suggested that the use of multiple-choice tests themselves may exhibit gender bias across different disciplines [71]. A study by Bolger aimed to explore the previously published idea that male students outperform females in multiple-choice tests and females perform more highly than males on open-response questions due to a superior verbal ability [71]. Results collected by analysing student performance on mathematics, Irish and English examinations showed a gender differences in favour of males on multiple-choice assessments for mathematics as well as languages.

1.2.4 Contextual bias of assessments

Although not examined explicitly in the research undertaken in this thesis, it is acknowledged that potential contextual bias of questions used in assessments can have an impact on student performance. Studies have noted that students may be sensitive to problem contexts depending on how familiar they are with the scenario described in the question [72, 73], with a review of existing research of context-based physics instruction conducted by Taasoobshirazi and Carr [74]. McCullough stated that conceptual tests such as the FCI contain questions which are heavily representative of stereotypically male contexts (for example rockets and cannonballs) and therefore may introduce a further gender dimension to the problem [75]. The author created an alternative version of the FCI, involving the same fundamental physics, but using extreme stereotypical female contexts and altering all situations to refer to female physicists in each question. Initial results suggested there was no significant change in female students' scores who took part in the test, but males taking the alternative version of the FCI had statistically significantly lower scores. However, it must be noted that female scores were close to the threshold score achieved by random guessing on a multiple choice test, therefore these results should be treated with some caution.

1.3 Attitudes to Science

The role played by students' attitudes to learning and studying physics on their attainment and participation in physics, and science in general, has been widely investigated and is recognised as a substantial factor in students' learning outcomes and can even influence their approaches towards studying their chosen subject [76, 77]. The attitudes and beliefs of students about physics may have a significant effect on their performance in the subject at university. A study by House found that self-belief and expectations can be a good predictor of exam performance [78].

As touched upon earlier in this chapter, the way students think about a subject may stem from their preconceptions of scientists and the study of science at a young age, gained through exposure to literature and media [31]. Studies have shown that high school pupils often have different perceptions of scientists and career opportunities depending on their gender [32]. Catsambis suggested that in secondary school girls may possess a more negative attitude towards science, despite performing at a higher level, and that males often consider it to be more applicable to future careers [79]. A paper by Osborne contains a comprehensive review of possible factors affecting the attitudes of pupils in STEM subjects and the implications these may have on students' performance [80].

A broad range of qualitative and quantitative techniques has been employed to measure students' interests and attitudes towards science. Traditionally this area has been explored using qualitative interviews and questioning of why students like particular subjects [80, 81]. Several survey instruments have been created to quantitatively assess students' beliefs. These include the Maryland Physics Expectation Survey (MPEX) [82], the Views about Science Survey (VASS) [83] and the Colorado Learning Attitudes about Science Survey (CLASS) [84].

The development of such instruments has allowed for students' attitudes to be quantised at different intervals throughout a period of study, thereby enabling the progression of these attitudes and beliefs to be observed.

The Colorado Learning Attitudes about Science Survey (CLASS) [84] is an instrument developed by the University of Colorado to measure the attitudes and beliefs of students about physics. The attitudinal survey has been extensively used across North America [84] and more recently worldwide [85]. It utilises a series of 42 statements marked on a 5 point Likert scale, in which students' beliefs are compared to those of physics 'experts'. The 'expert' response to each statement was validated through both surveys and interviews with 16 physicists, some involved in Physics Education Research, until consistent responses to all statements were established [84]. A more thorough discussion of the development and validation of the CLASS survey is presented in Chapter 2. In a paper published by the physics education research group at the University of Colorado, they distinguish between the ways in which experts and novices view physics, stating: "*Experts think about physics as a coherent framework of concepts which describe nature and are established by experiment. Novices see physics as isolated pieces of information that are handed down by authority (e.g. teacher) and have no connection to the real world, but must be memorized*" [86].

Extensive studies carried out by institutions in North America using the CLASS survey have observed changes in introductory physics students' level of expert-like thinking after a specified length of teaching [84]. Results showed that introductory physics students had differing views from 'expert' physicists, and it was generally found that they become less 'expert-like' over the course of an introductory course [84, 85]. This perhaps surprising decrease in measured attitudes and beliefs was seen at many institutions in different subject areas, both those that used traditional and innovative teaching methods [85, 87]. This reported decline has become an established fact in the area of Physics Education Research. There have, however, been published results showing an increase in students' attitudes when specific epistemologies have been targeted by changes in the curriculum or learning environments [88, 89].

The University of Colorado employed the CLASS survey to understand whether students' ideas about what physicists believed differed from their own opinions and whether these personal beliefs are affected by university instruction

[90]. The authors asked students to complete the survey twice, once answering the statements with respect to their personal beliefs and then again on how they think a physicist would respond. It was found that students' scores from the perspective of a physicist were significantly higher than those when answering the survey personally, suggesting that despite the fact that students know what physicists believe about learning physics, they are not in agreement. While students' personal scores showed an overall decrease after a semester of teaching, those from the viewpoint of an 'expert' remained relatively stable [90]. Overall, despite women scoring lower than males in their personal CLASS results, they showed a slightly higher perception of what they thought a physics 'expert' would believe. The reasons behind this observed decrease in students' attitudes towards science following instruction still remain unclear despite an increase in research into attitudes and beliefs across a broad range of subjects. It has been suggested that it may be linked to student confidence in their own abilities and understanding [91].

In Chapter 7 the topic of attitudes towards study will be discussed, with reference to the change in beliefs of first year students, as well as graduating students' views about their experiences of the undergraduate programme. The CLASS survey has been used both in order to investigate the attitudes in first year of study as was done in the original study by Adams *et al* [84], and to attempt to establish a UK measurement of academics' views of expert-like thinking.

1.4 Summary

This chapter has introduced several of the key issues surrounding gender disparity in undergraduate physics performance as well as assessment measures employed to determine the performance level and understanding of students. It is widely reported in the literature that females are greatly under-represented in STEM courses both at secondary and tertiary level, with research proposing that many factors may contribute to the relatively low percentage of women in university courses: stereotype threat, imposter syndrome, lack of role models and links between students' attitudes and their success. A large proportion of research surrounding gender physics performance has been undertaken at a secondary school level or within North American universities which have very different student demographics compared to UK institutions. This, alongside the fact that students' success in a course can impact whether they continue to pursue this major, makes it important to further investigate the factors that do and do not promote student learning in order to understand the gender gap and find ways to eliminate it. While some evidence of differential performance of male and female students has been documented, the gender differences in understanding of particular concepts has been less well understood or investigated and will be explored further in this thesis.

It is often asked why it is important to investigate gender differences in science education. Increasing the number of science students is imperative for the growth of industry and academia. Including a larger cross-section of the population, particularly females, may increase the interest of future students and ensure physics is more accessible and approachable to those who want to study the discipline. It is imperative that a student's interest is maintained as they move up the academic ladder. For instructors in particular, it is important to know the needs of the students at which a course is aimed and target areas in which the attainment is unbalanced. We need to be able to identify areas in which females are under-performing in order to help instructors to target difficulties students have with particular concepts for construction of future knowledge concepts. The results of such research can then be extrapolated to other disciplines.

1.5 Structure of the thesis

In addition to this introduction, this thesis contains a comprehensive description of the methodology and assessment tools used in this research which can be found in Chapter 2. Chapters 3-7 contain results collected throughout this study in relation to the topics introduced earlier, discussed alongside results form previously published studies. Chapter 3 focuses on results from a quantitative study of students' performance on conceptual understanding assessments at the University of Edinburgh as well as results from a comparison with the University of Hull and the University of Manchester. Chapter 4 comprises a question by question analysis from a Newtonian mechanics diagnostic test. In addition to this, student misconceptions are discussed in Chapter 5 through qualitative analysis carried out on questions of conceptual understanding. Differences found in male and female coursework and examination results, both in physics and other sciences, are discussed in Chapter 6. Chapter 7 looks at both student and academic attitudes and beliefs about learning and studying physics. Finally, Chapter 8 contains discussion and conclusions of this thesis research and suggestions for future work that could be carried out.

Chapter 2

Methodology

The focus of this chapter is to provide a comprehensive description of the educational context in which this study has been carried out as well as the different methods used in the analysis of collected data. Throughout this study different instruments have been implemented to investigate the existence of gender differences in students' performance and attitudes. This chapter serves as a reference, introducing the conceptual understanding diagnostic tests and assessment tools used. In addition to this, the statistical tests implemented to establish the significance of the data will be compared and discussed.

2.1 Education Context

The majority of the results presented in this thesis have been collected from students studying at the University of Edinburgh, with the exception of results used in a comparison with two English universities (University of Hull and University of Manchester), and results from UK members of the Institute of Physics which are discussed in later chapters. It is therefore important to establish the educational context and background of this institution as an environment in which students are exposed to different methods of teaching and learning.

The University of Edinburgh is a member of the Russell Group of UK universities focusing on academic research. Approximately 31,000 students are enrolled on either undergraduate or post-graduate courses across the university, of which almost 8,000 are within the College of Science and Engineering. Physics itself had 442 undergraduates and 178 postgraduate students in the 2012-13 academic year [92].

The School of Physics and Astronomy consists of four Institutes: The Institute for Condensed Matter and Complex Systems, The Institute for Particle and Nuclear Physics, The Institute for Astronomy and the Edinburgh Parallel Computing Centre. The school also has four research centres: the Centre for Science at Extreme Conditions, the Higgs Centre for Theoretical Physics, the UK Centre for Astrobiology and the Tait Institute. The wide range of research interests of staff working within the school is reflected in the variety of undergraduate degree programmes on offer to students. Students can enrol on one of ten degree programmes, five of which are based within the School of Physics and Astronomy and five of which are completed jointly with other university schools. Those offered by the School of Physics and Astronomy include Physics, Mathematical Physics, Computational Physics, Theoretical Physics and Astrophysics. Each of these provides students with the opportunity to study towards either a Bachelor of Science or to continue their studies for an additional year as a five year integrated Masters of Physics degree qualification. In order to progress onto the Masters programme students must first achieve a baseline of no less than 55% overall in their third year. The masters year(fifth year) is often viewed as a stepping stone for students considering studying a PhD following graduation and consists of a substantial independent research project. The five degrees offered jointly with other schools within the university are Mathematics and Physics, Physics and Music, Physics with Meteorology, Computer Science and Physics, and Chemical Physics. At the time of data collection for this thesis only Chemical Physics offered the option of a Masters degree course, with all other joint degrees ending with a Bachelor of Science qualification.

The majority of the research presented in this thesis has been conducted in first year introductory physics courses. These courses offer the opportunity to study students' levels of knowledge at the point of entry to university, prior to any university teaching. They have provided the chance to investigate the possible impact of teaching methodologies implemented by members of the Physics Education Research Group on learning. These teaching methods and the use of diagnostic testing have enabled changes in students' conceptual understanding and attitudes to be measured through data collection using the survey instruments discussed later in this chapter.

2.1.1 First year course structure

The Scottish Bachelor's degree has a typical duration of four years, with a first year that is slightly broader than that in England. The first year class studied here comprises both students for whom physics is a mandatory requirement for their degree programme, mainly students on physics degrees but may also include students on other degree programmes including Chemical Physics and Geophysics, and those who are taking it as an elective. In their first year of study students on the physics programme complete one third physics, one third mathematics (although delivered by the School of Physics and Astronomy) and one third a subject of their choosing.

The first year introductory physics course, 'Physics 1A', is an 11-week course with a typical class size of 200 to 300 students. Approximately half of the students enrolled on the course are studying with the intention of completing a physics degree and the other half are taking physics as a chosen elective. We have defined these students as 'majors' and 'non-majors' respectively for the purpose of this study. It is important to note that these are not necessarily equivalent to 'non-majors' at North American institutions, for which some comparable studies have been carried out, as all students taking our first year physics course are required to meet all entry qualifications for the course and thus differ only in that they have chosen not to pursue physics as their final degree subject. The number of 'non-major' students choosing to study physics in their second year is very small and all students in physics courses above this level are 'majors'.

This course has for many years been a focal point for curriculum innovation within the School, and details of the instructional design [93] and the role of studio-based workshop classes [94] have been reported elsewhere. Recently this course has included two further interventions: student-generated assessment content (PeerWise) [64, 95] and the introduction of the 'inverted' classroom approach [63, 96]. The move to an 'inverted' or 'flipped' classroom approach rather than a traditional lecture environment involves all students completing weekly reading assignments covering material which will be discussed in the upcoming week's lectures in addition to an online reading quiz completed prior to the Monday morning lecture. This reading quiz consists of five questions testing students' understanding of the content from the reading assignment and one question asking the students to comment on what they found difficult or most interesting about the reading material. The responses to the final question are then collated by the course instructor and used as basis for the focus of lectures. Students attend one hour lectures three times a week in addition to a three hour weekly tutorial workshop. Personal electronic voting devices ('clickers') are used during the lectures as an integral part of the inverted classroom approach to engage students in discussion about topics as well as provide the instructor with feedback on students' understanding. The lecturer uses clicker questions to promote discussion with the whole class, rather than a traditional lecture format in which the instructor presents the course content to the class in the format of a presentation. Routine gathering of data in first year physics classes, allows a baseline to be established to which any changes that occur during the teaching period can be compared. The use of in-lecture questions also provides the opportunity for data collection, some of the results of which are presented in Chapter 6 of this thesis.

2.1.2 Gender participation in physics

In order to examine gender issues in the undergraduate years of physics education, it is essential to first understand the demographics of our student population. On average, the physics undergraduate population at the University of Edinburgh comprises approximately 24% females, compared to 40% in the College of Science and Engineering and 56% across the entire university. Although low, this is higher than the UK national average for first degrees in physics, which has ranged between 18% and 21% since 2008 [21]. In this thesis, results were collected for students between 2006-13. Table 2.1 shows the number of students enrolled on the first year physics course ('Physics 1A') each year, as well as the gender proportions for each year's cohort. There was fluctuation in the number of students taking 'Physics 1A' each year, between approximately 200-300 students. The proportion of male students was consistently much higher than female students, on average 76%, reaching 82% in the 2012-13 academic year. Changes in the university recruitment method in 2010-11 have introduced a selection process rather than recruitment policy which resulted in an increase in the average entry qualifications.

Comparing this undergraduate gender profile with that of postgraduates and post-doctorates, we find that there exists little difference between populations. Of

Table 2.1: Number and proportions of male and female students completing the first year undergraduate 'Physics 1A' course.

Year	N Total	N Males	% Males	N Females	% Females
2006-07	239	179	74.9	60	25.1
2007-08	273	210	76.9	63	23.1
2008-09	291	224	77.0	67	23.0
2009-10	304	242	79.6	62	20.4
2010-11	207	146	70.5	61	29.5
2011-12	208	152	73.1	56	26.9
2012-13	295	241	81.7	54	18.3

approximately 150 PhD students in the School of Physics and Astronomy, 26% are female. Approximately 25% of post-doctorate researchers are female [97]. This is considerably higher than the UK national average of 17%. These figures suggest that we see no evidence of a leaky pipeline between PhD students and post-doctoral researchers, although the large number of postgraduates and post-doctorates that enter from other institutions means we need to be cautious about how we interpret the leaky pipeline with respect to the University of Edinburgh population.

Table 2.2: Number and proportions of major and non-major students enrolled in the first year undergraduate 'Physics 1A' course.

Year	N Total	N Majors	% Majors	N Non-Majors	% Non-Majors
2006-07	239	114	47.7	125	52.3
2007-08	273	108	39.6	165	60.4
2008-09	291	167	57.4	124	42.6
2009-10	304	157	51.6	147	48.4
2010-11	207	91	44.0	116	56.0
2011-12	208	89	42.8	119	57.2
2012-13	295	99	33.6	196	66.4

As mentioned previously, not all students enrolled on the first year introductory physics course are intending to complete a physics degree. The proportions of major and non-major students taking the introductory physics course in their first year of study are shown in Table 2.2. On average the number of students enrolled on a physics degree programme, approximately 45%, is outnumbered by the number of students taking it as an outside course. The proportion of majors and non-majors has shown large fluctuations between academic years. In particular, 2007-08 and 2008-09 showed an increase in the proportion of physics majors in the class. In comparison, 2012-13 had a similar number of students on the course, but a greatly reduced number of physics major students.

2.2 Pre- and post-test methodology

Integral to this study has been the use of pre- and post-instruction testing. The administration of a performance or conceptual based assessment prior to and after a period of instruction allows for the change in students' performance over a defined timescale to be measured. In many studies this occurs over one semester of teaching, but longer timescales, such as in longitudinal studies, can be examined. The pre-instruction assessment ('pre-test') acts as a baseline from which changes in a student's performance or attitude can be compared. Following a period of instruction or intervention, a post-instruction test ('post-test') is used and changes measured. Instructors must consider the timing of both of these tests, in particular that of the post-testing. It has been proposed that differences in results can be seen if students are tested immediately after the end of the instruction or if they are tested after a few days or weeks [98, 99]. In this thesis pre- and post-testing with the same students has been used in the administration of the Force Concept Inventory [46], which measures students' understanding of Newtonian mechanics, as well as in the use of the CLASS instrument, which measures the changes in student attitudes. It has also been used with the FMCE [36] and BEMA [49] diagnostic tests delivered to second year undergraduate physics students, discussed in Chapter 3.

Another method sometimes used in pre- and post-testing is the use of isomorphic testing. This involves students being given a different post-instruction test containing isomorphic questions relating to the same topic or concepts tested in the pre-instruction test. It has been argued that students who are able to make connections between two different problems with different surface features gain the additional skill of being able to transfer their knowledge and therefore become more expert-like in their problem solving ability [100]. The use of isomorphic questions may also avoid potential concerns that the gains witnessed in student performance may be due to students remembering surface features of questions when they did the pre-test. Isomorphic questions were not used in this study. Any pre- and post- testing presented in this thesis was undertaken using the same instrument prior to and post instruction to ensure consistency between assessments and year group comparisons.

2.3 Tests of conceptual understanding

Diagnostic tests are frequently used in Physics Education Research to measure the change in students' understanding over a period of time. They can be employed as an instrument for assessing the effects of new teaching methodologies or interventions. Concept inventories are typically administered as multiple-choice assessments to ensure the reproducibility of results across large cohorts and between institutions. In many cases distractors used for each question have been specifically chosen to highlight a specific idea or misconception held by students. These misconceptions are often determined through interviews with students or through open-response questions.

During their development the reliability and validity of the concept inventory must be checked. A test is considered valid "*if the skills or knowledge it measures are directly relevant to the stated domain of the test. Validity cannot be assessed statistically and is usually determined by a consensus of expert opinions*" [101]. The validity of an instrument can be explored through three main areas: content validity, construct validity and criterion validity. The content validity measures the degree to which the instrument reliably represents the content it is created to test. Construct validity refers to the extent to which the test measures what it purports to be measuring. Criterion validity compares the outcome of the test with that of other validated measures. The use of such validation techniques will be discussed with respect to each individual concept inventory presented in this chapter.

The reliability of each individual test item must be established alongside the reliability of the test as a whole. A test can be considered to be reliable if "*one can have confidence that the same students would get the same score if they took the test more than once. In addition, on a reliable test, a large fraction of the variance in scores is caused by systematic variation in the population of* *test takers; students whose levels of understanding or mastery are different will achieve different scores on the test*" [101]. The reliability of an individual test item can be determined using several statistical tests. A few of those which were used in the original construction of the concept tests used in this thesis are discussed below.

The difficulty level of a test item can be calculated using the **Item Difficulty Index (*P*)**. This is a ratio of the number of correct responses on the test item ($N_{correct}$) to the total number of question responses (N). A higher P value indicates that a higher proportion of students answered correctly. The Item Difficulty Index ranges from 0 (no one answering correctly) to 1 (everyone answering correctly), with an optimum value of 0.5 [101]. Naturally, there will be some variation in the difficulty of test items, and in many cases an averaged difficulty index is used to determine the overall test difficulty ($\bar{P}$).

$$P = \frac{N_{correct}}{N}$$

The **Item Discrimination Index (*D*)** can be employed to determine the ability of test item to discriminate between students with a strong understanding of the material and those with less robust knowledge. The discrimination index is calculated by counting the number of correct responses by students above the median total score (N_H) and similarly counting the number of correct responses by those students below the median total score (N_L). It ranges from -1 (all students in the lowest performing group answer correctly but no one in the top performing group is correct) to +1 (all students in the top performing group answer correctly and no one in the bottom performing group is correct).The discrimination index is then calculated as

$$D = \frac{N_H - N_L}{N/2}$$

where N is the total number of responses. Similarly, D can be calculated using the top 25% of student responses and the bottom 25% of student responses [101]. This method eliminates the middle 50% of data and includes only the most consistent quartiles, thereby reducing the possibility of underestimating the discrimination index.

$$D = \frac{N_{top25\%} - N_{bottom25\%}}{N/4}$$

A **Point Biserial Coefficient** (r) measures the correlation between a respondent's score on an individual test item and their overall test score and ranges between -1 and +1. Ideally all the test items on the concept inventory should have a high correlation with the final test score. A high r value indicates that a student who answers correctly on a test item is likely to achieve a high overall test score. When calculating r for a particular test item the Item Difficulty Index (P) is used along with the average overall test score of those who answered that item correctly ($\bar{X}_c$), the average overall test score for the entire sample ($\bar{X}$) and the standard deviation of the total sample scores (S_x).

$$r = \frac{\bar{X}_c - \bar{X}}{S_x}\sqrt{\frac{P}{1-P}}$$

The **Ferguson's Delta** (δ) is another example of a test statistic which looks at how respondents' scores are distributed over the total possible range of scores. For a well designed test that has been created to distinguish between students' levels of understanding it is expected that a wide range of final scores is seen. The Ferguson's Delta takes a value ranging between 0 and 1, with values greater than 0.9 deemed to show good discrimination [101]. It is calculated by comparing the final scores of two respondents and determining the number of equal and unequal scores within the sample. The Ferguson's Delta is calculated as

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - N^2/(K+1)}$$

where N is the total number of students, K is the number of test items and f_i is the number of occurrences of each score.

A wide variety of multiple-choice diagnostic tests are available to test conceptual understanding in different subject areas [48]. In this section three diagnostic tests in Newtonian mechanics and electricity and magnetism will be discussed, along with a brief overview of the validation processes used in their design.

2.3.1 Force Concept Inventory (FCI)

One of the most extensively used assessment instruments is the Force Concept Inventory (FCI) which was developed by Hestenes *et al* as a tool for measuring students' understanding of Newtonian mechanics [46]. The creation of this test originated from the fact that students' commonsense beliefs about force and motion are often found to be incompatible with Newtonian mechanics concepts [52]. A full version of the FCI diagnostic test can be found in Appendix A. The survey has been employed extensively in institutions throughout North America and more recently worldwide. The authors highlighted the three main applications of the survey instruments. Firstly it can be used as a diagnostic tool for assessing students' misconceptions and to bring these to the attention of a course instructor so that interventions can take place to instigate conceptual change. Secondly, the FCI can be use as a measure of the effectiveness of instruction. Administering this test both at the start and completion of a specific course (as a pre-test and post-test) allows for it to be used as a measure of how a specific course has affected students' understanding of tested concepts. Finally, the authors suggest that it could be used as a placement exam in conjunction with the Mechanics Baseline test [102].

One of the key features of this instrument is that students are forced to make a choice between commonsense misconceptions and Newtonian concepts when choosing a multiple-choice answer. A 'common sense knowledge state' is defined as knowledge originating from a person's personal experiences and not derived from formal physics instruction. The survey authors reported that the concept inventory "*is not a test of intelligence; it is a probe of belief systems*" and that incorrect responses are often more informative than correct answers as they inform instructors of students' preconceived beliefs. Halloun *et al* created a detailed taxonomy of the misconceptions probed by the FCI [103]. These misconceptions were grouped into six categories: Kinematics, Impetus, Active Force, Action/Reaction Pairs, Concatenation of Influences and Other Influences on Motion (Resistance and Gravity). The original FCI consisted of 29 questions which probed six individual concepts relating to Newtonian mechanics: Kinematics, Newton's First Law, Newton's Second Law, Newton's Third Law, Superposition Principle and Kinds of Force. Each inventory question was associated with one of these concepts, with the exception of question 12 which discusses both gravity and buoyancy (See Appendix A).

Validation of the Force Concept Inventory

The Mechanics Diagnostic Test (MDT) was originally administered as an open-ended qualitative assessment to more than 1000 college physics students. Students' responses to each of the questions were then used to develop multiple-choice answers probing the most common misconceptions by students [103]. The content and face validity of the test were established by presenting the test to both graduate students and physics faculty. Face validity looks at the extent to which an assessment is viewed as covering the content or aims it has been created to measure. In addition to this, introductory-level physics students were interviewed to establish whether each question and its possible answers were correctly understood. Students were asked to reiterate their answers from their written test and did so "*virtually without exception*", and therefore the reliability of results was seen through the confirmation that students' responses to the open-answer and multiple-choice tests were highly comparable. The authors found that test results indicated that the "*students' answers reflected stable beliefs rather than tentative, random, or flippant responses*".

The FCI is heavily based on the Mechanics Diagnostic Test [52], with about half the questions from the original version replicated from this assessment. The reliability and validity of the FCI was not however fully re-established. Twenty students were interviewed about their responses to the inventory items. Further interviews with 16 graduate students highlighted a clear lack of understanding of Newton's Third Law, with several students unable to recognise the situations in which it could be applied or draw a suitable free-body diagram. Two test items were eliminated because it was found that many students had difficulty understanding the wording of these questions. Test scores on the MDT and FCI showed similar scores for similar student populations. A revised version of the

FCI contains 30 test items. It was noted that this version has *"fewer ambiguities and a smaller likelihood of false positives"* [44].

Hestenes *et al* issued this test to over 1,500 secondary school students and 500 university students and found that students from all participating institutions had unexpectedly low pre-test scores. They went on to classify a score of 60% on the FCI as an 'entry threshold' for Newtonian mechanics [46]. Below this score Hestenes *et al* proposed that students have an insufficient understanding of Newtonian concepts for effective problem solving. A follow up to this research set a score of 85% on the FCI as a 'mastery threshold', where a student was classed as having a full grasp of Newtonian mechanics [104].

In this thesis a revised and extended version of the FCI containing 33 questions was used in the first year introductory physics course between 2006-10 [105]. A copy of this test can be seen in Appendix B. After this time, we used the revised version of the FCI containing 30 questions [46]. This enabled our results to be more readily compared to those from other institutions both in the UK and in North America. Quantitative results collected from first year students at the University of Edinburgh and qualitative results from interviews with students both here and in other UK universities are discussed in Chapters 3, 4 and 5 of this thesis.

2.3.2 Force and Motion Conceptual Evaluation (FMCE)

The Force and Motion Conceptual Evaluation (FMCE) is a conceptual survey containing 47 multiple-choice questions probing understanding of Newtonian mechanics [36]. These questions are arranged in five categories: Newton's First and Second Law, Newton's Third Law, Velocity, Vertical Motion and Energy. The FMCE uses a combination of graphical and verbal representations of one dimensional force and motion. A full version of the FMCE can be seen in Appendix C. As with the FCI, this assessment is traditionally used as a pre- and post-instruction measure of the change in students' conceptual understanding.

The multiple-choice answer options were established through free-response answers and student interviews. Statistical tests of reliability were not formally established by Thornton and Sokoloff, but instead an investigation into the correlation between a student's written responses and their multiple-choice answers was completed. A later study by Ramlo found that factor analysis and content validity showed the assessment to be a reliable instrument for measuring students' understanding of force and motion [106].

A further study by Thornton *et al* compared test results from the FMCE and the FCI [107]. They found that, although there was a very strong correlation between average score on the two assessments, students typically had higher scores on the FCI. They commented that the FMCE "*provides a more detailed measure of student understanding by virtue of a greater number of items covering a narrower range of topics*" whereas the FCI "*is a good indication of student's ability to solve problems dealing with Newtonian mechanics*" [46]. In this thesis the FMCE was used to test the initial conceptual knowledge of second year undergraduate physics students at the University of Edinburgh. Students had previously completed a pre- and post-test FCI in their first year and therefore would have been familiar with the concepts tested by the instrument. Results from the use of this survey are discussed in detail in Chapter 3.

2.3.3 Brief Electricity and Magnetism Assessment (BEMA)

The Brief Electricity and Magnetism Assessment was developed by Chabay and Sherwood to measure understanding of introductory concepts of electricity and magnetism for calculus-based physics courses [49]. This instrument contains 31 multiple-choice test questions. A copy of this instrument can be seen in Appendix D. It is not based around a specific curriculum but instead focuses on topics which are typically covered in introductory calculus-based electricity and magnetism courses and matter and interaction curricula. The majority of items are qualitative, with a few quantitative questions involving only simple mathematical calculations.

The validity and reliability of BEMA was explored through statistical tests on item analysis and whole test reliability. When establishing the content and face validity of the original assessment, the test was checked by eight faculty members at Carnegie Mellon University who had taught on an electricity and magnetism course within the last five years. The faculty members checked whether questions dealt with topics which had been covered in the course on which they had taught. Any questions that had not been covered by lecturers were eliminated from the assessment. Subsequent testing was undertaken with a group of senior physics majors. An earlier version of the test contained both multiple choice questions and short-answer quantitative questions. These short-answer questions were changed to multiple-choice questions in the final version of BEMA. The reliability and discriminatory power of individual test items were evaluated using five different statistical tests including the item difficulty index and item discrimination. Full details of the results of these tests can be found in the paper by Ding *et al* [101]. Results from the use of this conceptual test will be presented in Chapter 3 of this thesis.

2.4 The Colorado Learning Attitudes about Science Survey (CLASS)

This section aims to provide an overview of the key features of the CLASS survey instrument. The CLASS instrument was first developed at the University of Colorado by Adams *et al* in order to measure changes in students' attitudes and beliefs about physics over a semester [84]. A full copy of this survey can be found in Appendix E. It was developed as an extended and more extensive version of the Maryland Physics Expectation Survey (MPEX) [82] and the View about Science Survey (VASS) [83]. It was first presented at the 2004 Physics Education Research Conference (PERC) [108]. Unlike conceptual understanding tests, the CLASS survey was created in order to provide a survey that specifically probed individuals' opinions and investigated how students thought about physics and approaching physics problems. The authors stated that the statements were written specifically to have a single interpretation and be suitable for use with students over a range of ability levels and physics courses. The survey itself has been subsequently modified for use in other science disciplines including Chemistry [109] and Biology [110].

The attitudinal survey contains a series of 42 statements to which students rate their level of agreement on a five-point Likert scale from strongly agree to strongly disagree. Student scores on this assessment can be analysed both as an overall agreement score and on individual categories of questions. These categories were determined through empirical groupings based on responses from students, rather than groupings determined by survey creators based on characterisations of student beliefs. Factor analysis was carried out at the time of the survey's design to determine the series of categories in which questions could be grouped. In total eight categories were defined; *Personal Interest*, *Real World Connection*, *Problem Solving General*, *Problem Solving Confidence*, *Problem Solving Sophistication*, *Sense Making and Effort*, *Conceptual Understanding* and *Applied Conceptual Understanding*.

Students' responses to the survey are directly compared to those of physics 'experts' and are therefore given a percentage of 'expert-like' thinking score. This allows for students' attitudes to be compared to those of physics academics and also enable students for various cohorts and institutions to be compared to a specific reference point. The validation of the expert responses is discussed in the next section.

2.4.1 Survey design and validation

As stated above, the survey statements were originally based on those from the MPEX survey instrument created by the Redish Group at the University of Maryland [82]. In order to modify some of these statements to fit the guidelines of the survey, interviews were carried out with both students and experts [84]. This allowed for the statement wording to be such that it represented the vocabulary used by students and therefore eliminated any ambiguity in the meaning of each statement. For example, the authors stated that the word "domain" and "concepts" were not prevalent in students' explanations and should therefore be avoided to make the survey accessible to a wide range of students. Similarly, it was decided that the survey should not refer to specific courses to allow for it to be universally applicable. In addition to this, new statements referring to problem solving, personal interest and sense making and effort were created.

The survey contains a fail-safe question which states "*We use this statement to discard the survey of people who are not reading the questions. Please select agree - option 4 (not strongly agree) for this question to preserve your answers*". This statement allows for unreliable data to be omitted from further analysis, although it does not necessarily eliminate survey results from students not taking the survey seriously. For a student's responses to be included in scoring, a minimum of 32 of the 36 statements that are scored must be completed.

Students' responses are scored by determining a percentage favourable agreement (percentage to which the student's response is in agreement with the 'expert' response) and percentage unfavourable agreement (percentage to which the student's response disagrees with the 'expert' response). Survey results are given an 'overall' percentage favourable score as well as individual scores for the eight categories mentioned above. The survey itself contains 42 attitudinal statements, 27 of which have been included in the eight categories. A further 9 statements are included in the 'overall' percentage score. The 6 remaining statements do not have a defined 'expert' response and "*are statements that are not useful in their current form*" [84]. The survey authors have not offered further explanation of why these statements have nevertheless been included in the final instrument.

Answers to each of the statements are scored on a five-point Likert scale (from Strongly Agree to Strongly Disagree). The authors commented that interviews with students highlighted that those students who chose the neutral answer to a statement did so because they did not know how to answer the questions, had no strong opinion, had different opinions from different physics courses they had experienced, or were unsure whether to answer "*what they think they should do versus what they actually do in practice*". Because of this, the five-point scale was treated as an ordinal scale, in which the difference between each possible response is not an equal distance. Despite this, the five-point scale was collapsed into a three-point scale (agree, neutral, disagree) for analysis. The authors stated that students felt "*that agree vs strongly agree (and disagree vs strongly disagree) was an important distinction and that without the two levels of agree and disagree they would have chosen neutral more often*".

The survey was validated using four different validation processes: face validity through interviews and survey responses from both students and physics faculty; construct validity with survey responses collected and analysed from thousands of students from which statement categories were determined; predictive validity from correlation of students' incoming course performance and beliefs; and concurrent validity comparing responses to expert results.

One key feature of the CLASS survey is that a student's attitudes and beliefs are directly compared to those of physics experts. The predefined expert responses to each statement were determined through a series of interviews with physics faculty members [84]. Initially three experts provided comments on statements which could be interpreted in multiple ways. After this, a further 16 experts took the survey and their answers compiled to give the 'expert view' used in the survey scoring. Each of these experts was a member of the physics faculty at the University of Colorado and had prior experience in teaching introductory physics courses. Several of these faculty members were also involved in physics education research, and all were male. Where there was not a general consensus between the staff responses, discussion was used to try and agree on an expert response. Their responses were mutually consistent for all but four of the CLASS statements. These four statements were subsequently not included in any of the eight categories and focused on students' learning styles and beliefs about the nature of science. The reliability of this expert response is investigated from a UK perspective in Chapter 7 of this thesis.

The survey was further validated by conducting interviews with 34 students across six different physics courses, including both major and non-major students as well as students representing both genders and different ethnic backgrounds [84]. Students first completed the survey before being interviewed about how they interpreted each statement. One important consideration that the authors explored was whether students answer the survey thinking about what they themselves think about physics or whether they answer the statements in the way they believe a physics expert should respond. This was investigated in a study by Adams *et al* in which students were explicitly asked to answer each question twice, once as themselves and once as a physics expert [111]. Results showed that students did make a distinction between the two viewpoints. It is therefore important, when administering this instrument, to make it clear to students that they should answer the statements from their own viewpoint.

The CLASS survey was implemented in first year undergraduate courses at the University of Edinburgh using pre- and post-test methodology. This enabled any changes in students' attitudes and beliefs towards physics to be measured over a specific period of time. Further details of how it was implemented and subsequent results are discussed in detail in Chapter 7 of this thesis.

2.5 Statistical Tests

Quantitative data collected in this thesis has been analysed using tests of statistical significance to determine whether observed changes in a cohort's performance or attitudes are evidence of real changes, or simply numerical fluctuations. Tests were used to compare both pre- and post-test results of a single cohort and to compare results from different populations. Data analysis in this study has been conducted using Microsoft Excel and SPSS software, both from integrated statistical analysis features and raw data calculations. Each of the statistical tests used in data analysis in this project will be presented and its uses discussed.

2.5.1 T-tests

A **T-test** is a parametric test used to determine whether the mean values of two distributions show a statistical difference, where the underlying populations are normally distributed. The t-test assumes a null hypothesis that the two distributions are the same and that experimental manipulations have no effect on the populations. If the null hypothesis is rejected, it can be assumed that the means of the two samples are different because of different experimental manipulation. In the case of results presented in this thesis, it is not the the result of randomised controlled experiments, but an influence or relationship between the variables being measured.

The test statistic can be calculated from

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_d\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\bar{x}_1$ and $\bar{x}_2$ refer to the means of the first and second sample and n_1 and n_2 are the sample sizes of the two populations. The standard deviation of the results of the two samples is denoted by s_d. In all results presented in this thesis a confidence level of 95% has been assumed. Therefore, if the t-test results in a p-value of $p \leq 0.05$ there is said to exist a statistically significant difference between the means of the two samples.

When using a t-test one must also determine whether to use a one or two-tailed t-test. The number of tails specifies the predicted direction of the measured difference between two groups. If there is no predicted direction made during data collection or analysis, then a two-tailed t-test is used. If the hypothesis states that the two data sets will differ from each other in a certain direction, then a one-tailed t-test is used. In all cases presented in this thesis a two-tailed t-test has been used.

There are two different types of t-tests that can be used depending on whether the two distributions being compared comprise the same or different populations [112]. The first of these is a 'Paired t-test' ('Dependent t-test'). A 'Dependent t-test' is used if the same population is tested over two different variables and assumes that the distribution of the differences in the scores is normal. For example, such a test is used when comparing the mean pre-test score of the male population to the mean post-test score of the same male population. The SPSS output from a dependent t-test provides both the Pearson's r value and the two-tailed significance. The test statistic t is calculated by dividing the differences of the means of the two samples by the standard error of the differences. SPSS uses the degree of freedom to determine the probability that this t value could be obtained if the null hypothesis was true and denotes this as the two-tailed significance value p [112].

An 'Independent t-test' is used to assess data where two different populations are tested against a fixed variable. For example, this is the case when comparing the mean pre-test score of a male population to the mean pre-test score of a female population. Unlike for the Dependent t-test, there exist two possible output options for an Independent t-test: equal variances assumed or equal variances not assumed. The Levene's test is used to test whether the assumption that the variances in the two groups are equal holds true. If the Levene's test shows a significance of $p \leq 0.05$, then the variances are assumed to be unequal. The t-statistic is calculated by dividing the difference between the means of the samples by the standard error of the sampling distribution. For an independent t-test the number of degrees of freedom is calculated by adding the sample sizes of the two populations together and subtracting the number of samples. Once again the two-tailed significance value is noted p.

2.5.2 ANOVA

The **Analysis of Variance (ANOVA)** test is used to test whether there exist any statistically significant differences between the means of three or more independent samples [112]. The ANOVA assumes that the distribution within the sample groups is normally distributed. Completing multiple t-tests could result in an increased chance of a 'false positive' result, showing statistically significant differences between groups which do not exist. As the number of t-tests conducted between samples (n) increases, the probability of finding a wrongly significant independent result will also increase. This error arising from multiple statistical tests on the same data is referred to as the familywise error.

$$Error = 1 - (0.95)^n$$

The F-statistic produced by the ANOVA compares the amount of systematic variance to the amount of unsystematic variance [112]. The output of the ANOVA test statistic cannot specify which samples are significantly different from each other, but only indicates that at least two samples are different. Post hoc tests can then be used to determine which groups show statistically significant differences from one another. A 'one-way' or 'two-way' ANOVA can be used depending on the number of independent variables.

2.5.3 Mann-Whitney U test

It is important to note if the data collected is not normally distributed. It is possible to rely on the central limit theorem which states that if the sample size is sufficiently large, the distribution of the sample can be considered to be normal. Alternatively, a non-parametric test can be used. The **Mann-Whitney U** test is a non-parametric t-test used to determine whether significant differences exist between two independent groups when the dependent variable is ordinal or continuous, but not normally distributed [112]. For the data included in this thesis which did not demonstrate a normal distribution, the Mann-Whitney U test was used. For example, scores on post-instruction FCI tests were not normally distributed since scores on the post-test were close to the maximum, with little fluctuation above the mean value. The Mann-Whitney U statistic compares ranked data for each condition, with the U statistics denoting the difference between the two rank totals, taking into account the different sample size. The smaller the U value, the less likely it is that the measured difference has occurred by chance.

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

where n_1 and n_2 are the sample sizes of groups 1 and 2 and R_1 is the sum of the ranks of group 2.

2.5.4 Kruskal-Wallis Test

The **Kruskal-Wallis** one-way analysis of variance test is used to determine if the differences between three or more samples is significant by testing the null hypothesis that all the populations have identical distribution functions. Unlike the ANOVA test, it does not assume a normal distribution [112]. If the test shows significance, a difference exists between at least two of the samples and a further test must be used to determine which of the samples are significantly different. As with the Mann-Whitney U test, the Kruskal-Wallis test is based on ranked data. The test statistic is given by

$$H = \frac{12}{N-(N-1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

where N is the total sample size, n_i is the sample size of a specific group and R_i is the sum of the ranks of each group.

2.5.5 Chi-squared test

The Pearson's chi-squared test can be used to test whether the observed frequency distribution of a sample diverges statistically from the expected values under the null hypothesis of no association. The expected values calculated for each cell (i, j) in a two-way table d_{ij} are equal to

$$\bar{d}_{ij} = \frac{(\sum_{i'} d_{i'j})(\sum_{j'} d_{ij'})}{n}$$

where n is the total number of observations in the table. The chi-square test compares the distributions of the variable rather than just the mean of the distribution and is calculated using the equation shown below.

$$\chi^2 = \frac{\sum_{ij}(d_{ij} - \bar{d}_{ij})^2}{\bar{d}_{ij}}$$

This is distributed as a χ^2-distribution with $(N_{column} - 1)(N_{row} - 1)$ degrees of freedom, and can test whether the observed value for the statistic is significant by looking at the p value for the χ^2 distribution. It has been used in this thesis to compare the distributions of two samples to determine if they could be derived from the same population.

Chapter 3

Conceptual Understanding in Undergraduate Physics

As was explained in the introductory chapter of this thesis, there are a number of factors influencing male and female academic performance in physics and other sciences at undergraduate level. There is evidence to suggest that the pedagogy of science classes can improve student performance and may result in a narrowing of the observed gap between male and female attainment [44, 50, 51, 55, 62]. A study conducted by Lorenzo *et al* during an introductory calculus-based physics course at Harvard University showed that both male and female learning gains increased after one semester following the instigation of interactive engagement techniques and Peer-Instruction [55]. The pre-instruction gender gap was fully closed at the end of the semester. This is not, however, a result that has been universally reproduced and there is a suggestion that both instructor effects and the background experience of students may be factors in explaining observed differences [56]. We are motivated to re-examine the gender performance difference in the context of UK physics undergraduates, in part because of the lack of clear consensus in the literature, but more so because both the format of university education and the education background prior to coming to university are very different from those in the United States. In addition to this, many of the studies that have been carried out at US institutions relate to courses delivered to a cohort of students made up primarily of non-majors in physics. These courses often have a gender profile atypical of those at UK institutions.

This chapter focuses on the measurement of students' conceptual understanding in the first two years of university and on exploring whether students at the University of Edinburgh show evidence of a gender performance discrepancy. Students' understanding of concepts of force and motion, as well as electricity and magnetism, were measured using diagnostic tests. The change in performance of first year undergraduate students at the University of Edinburgh will be discussed with reference to data collected over seven consecutive years of conceptual testing of Newtonian mechanics. First, the performance of the class as a whole is reviewed, both prior to and after instruction. An in-depth study of the quantitative changes in male and female conceptual understanding measured over a period of instruction will also be presented, followed by conclusions from a study comparing observed gender discrepancies at three UK universities.

The first year of study at Edinburgh allows students from other degree disciplines to choose to enrol on first year physics courses as an elective. This means that a high percentage of our first year physics students are not intending to continue on a physics degree programme, despite holding similar school-leaving qualifications to those intending to graduate from the School of Physics and Astronomy. In this chapter the differences seen between these students and those on the physics degree programme are compared. When comparing results to those from previously published studies by institutions in North America, it must be acknowledged that their 'non-majors' represent a very different cohort of students. Whereas at the University of Edinburgh all students taking 'Physics 1A', including 'non-major' students taking it as a chosen elective, are required to possess the entry requirement physics qualification, for students taking introductory physics courses at US institutions a final year school qualification in physics is not always essential. In many cases there is also a large contrast in the number of physics majors in US and UK introductory physics courses. For example in a study by Kost *et al*, they commented that only 6% of their population were declared physics majors, whereas approximately 50% of the first year cohort at Edinburgh are majors [3].

In order to address whether observed gender differences are unique to first year studies or to the understanding of Newtonian mechanics, student performance on conceptual understanding assessments in second year physics courses, focusing on mechanics and electricity and magnetism, have been measured. The make up of the second year cohort differs from that of first year in that almost all students taking the second year course are enrolled on a physics degree with only a small number of students taking it as an additional course.

3.1 Conceptual understanding in first year undergraduate physics

This section discusses a study of the change in the understanding of Newtonian concepts of force and motion of the first year physics class at the University of Edinburgh. When investigating students' conceptual understanding it is necessary to establish a foundation from where changes in performance can be measured. As well as establishing the improvement of the class as a whole, this study aimed to investigate the existence and extent of a possible performance gender gap in the first year introductory physics course. Results presented in the following sections, from a typical UK cohort of students in which females are under-represented in the physics population, focus on changes in students' performance on conceptual understanding tests after one semester of teaching. In addition to evaluating differences between male and female students, it was examined whether students enrolled on a physics degree programme performed significantly differently to those who chose to study it as an elective.

3.1.1 Educational background of cohort

The first year physics cohort at the University of Edinburgh comprises students entering from a variety of different educational backgrounds. Approximately 55% of students enter from Scottish schools, 20% from across the rest of the UK and 25% are international students [113]. The difference in backgrounds means that students are starting the course with a wide variety of different school leaving qualifications: Highers, Advanced Highers, A-Levels and a small number with Irish Leaving Certificate or the International Baccalaureate. Students from a Scottish educational background typically enter with either Higher or Advanced Higher qualifications awarded by the Scottish Qualification Authority (SQA). Students typically study five subjects at Higher level before proceeding to tertiary education, although the majority of pupils remain in secondary education to complete Advanced Highers in some of these subjects. Pupils from the rest of the

UK (and in some independent schools in Scotland) complete Advanced Subsidiary Levels (AS-Levels) and Advanced Levels (A-Levels) in their final years of school. It is common for students to complete four or five AS-Levels in their penultimate year, proceeding to choose three subjects to continue to A-Level. The format of the A-level curriculum means that students are taught six modules over the course of their two years of study. However the course modules for A-level subjects differ depending on the exam board used in a particular school or region. Because of the variation in the physics and mathematics educational background of incoming students, including the content, assessment type and practical skills to which they have been exposed, it must be expected that there exists some variation in students' prior learning, despite entering with broadly equivalent qualifications.

3.1.2 Methodology

As described in Chapter 2, the Force Concept Inventory (FCI) is one of the most extensively used diagnostic tests of conceptual understanding in physics and has served as a benchmark for the creation of a wide variety of instruments and inventories in science education research [46]. The FCI is conventionally administered to students twice, prior to instruction and again after the completion of the course (pre- and post-test methodology).

In the 2011-12 academic year the decision was made to begin using the original version of the FCI in place of the alternative slightly extended version (FCI_{ext}) that had been in place since 2006 [105]. Although this does not allow a direct comparison with previous years' data from the University of Edinburgh, it enables us to directly compare overall results and responses to individual questions with both the literature and other institutions. FCI data has been collected continually at the University of Edinburgh in the introductory physics course ('Physics 1A') for the past seven years between 2006-13, although when introduced in 2006 the FCI_{ext} was not fully embedded in the curriculum as part of a weekly assessment. Pre- and post-test methodology was employed with the higher mark of the students' two attempts contributing 3% to their final course mark (equivalent to one weekly coursework assessment). In addition to a change in the assessment instrument, in 2010-11 active recruitment of physics undergraduate students was replaced by a selection process which may have had an effect on the incoming cohort of students, with higher entrance qualifications being introduced.

The test is delivered online through the course VLE (Virtual Learning Environment) system, with students given a time constraint of 90 minutes in which to complete the test. The test remains open for students to complete at any time over a week long period. Although students are encouraged to complete the questions in the order in which they are presented, there is no restriction on the order in which they are able to answer them before submitting the test.

The FCI consists of 30 questions which probe six individual concepts relating to Newtonian mechanics: 'Kinematics', 'Newton's First Law', 'Newton's Second Law', 'Newton's Third Law', 'Superposition Principle' and 'Kinds of Forces'. A complete version of the FCI can be found in Appendix A. In answering these questions students are compelled to choose between Newtonian concepts of force and common misconceptions. Hestenes *et al* issued this test to over 1,500 secondary school students and 500 university students and found that students from all participating institutions had unexpectedly low pre-test scores [46]. As mentioned in the previous chapter, the authors classified a score of 60% on the FCI as an 'entry threshold' for Newtonian mechanics. Further research set a score of 85% on the FCI as a 'mastery threshold', which they believe indicates that a student should have a full grasp of Newtonian mechanics [104]. These classifications have allowed us a benchmark upon which to judge our students' conceptual understanding at the point of entry. The FCI_{ext} administered between 2006-10 contains 33 questions of which 19 are consistent with the original version of the FCI: testing the same concept, but with minor changes to wording or representation. A copy of the FCI_{ext} can be found in Appendix B.

3.1.3 Simpson's Paradox

The Simpson's Paradox is a situation in which a trend or a statistical significance between different sample groups disappears or is reversed when these samples are combined into one data set. Similarly, a statistical significance may appear when the data is broken down into different sample groups. It is particularly important to consider this paradox when forming conclusions from individual data sets which have been aggregated.

For example, if we observe male and female students' scores in two different year groups it may be found that males had a slightly higher score than females in both cases. However if the data for each gender is combined across year groups, as shown by the hypothetical data in the table below, a different conclusion could be made, with females showing a marginally higher average score than males.

	Gender	N	Average Score
Year 1	Females	20	0.30
	Males	30	0.31
Year 2	Females	30	0.40
	Males	20	0.41

	Gender	N	Weighted Average Score
Years 1 and 2	Females	50	0.36
Years 1 and 2	Males	50	0.35

The implications of the Simpson's Paradox is particularly relevant when discussing results of male and female physics majors and non-majors in section 3.1.8, in which small number statistics show different results between individual year groups than in the combined data sets.

3.1.4 Whole class FCI performance

In this section the results collected from the FCI will be presented with reference to the performance of the whole first year cohort. In this and all following results only matched data points have been included in analysis; i.e. only students who completed both the pre- and post-test FCI. As a result of this, the sample size for each cohort is lower than the total number of enrolled students on the course.

The 'Physics 1A' course covers all the material tested in the FCI within the first five or six weeks of the semester, with the remaining weeks focusing on energy conservation, momentum and simple harmonic motion. Students are tested on their knowledge of Newtonian mechanics both in weekly coursework assignments and a midterm class test which comprises a past exam question. The FCI pre-test is completed in week one of the first semester. The post-test is then distributed at the end of the semester, approximately five weeks after formal teaching of these topics has been completed.

Between 2006-10 the content of the 'Physics 1A' course, as well as the teaching methods used, remained unchanged. A Kruskal-Wallis test was carried out on the data to look for any statistically significant differences between the five sets of

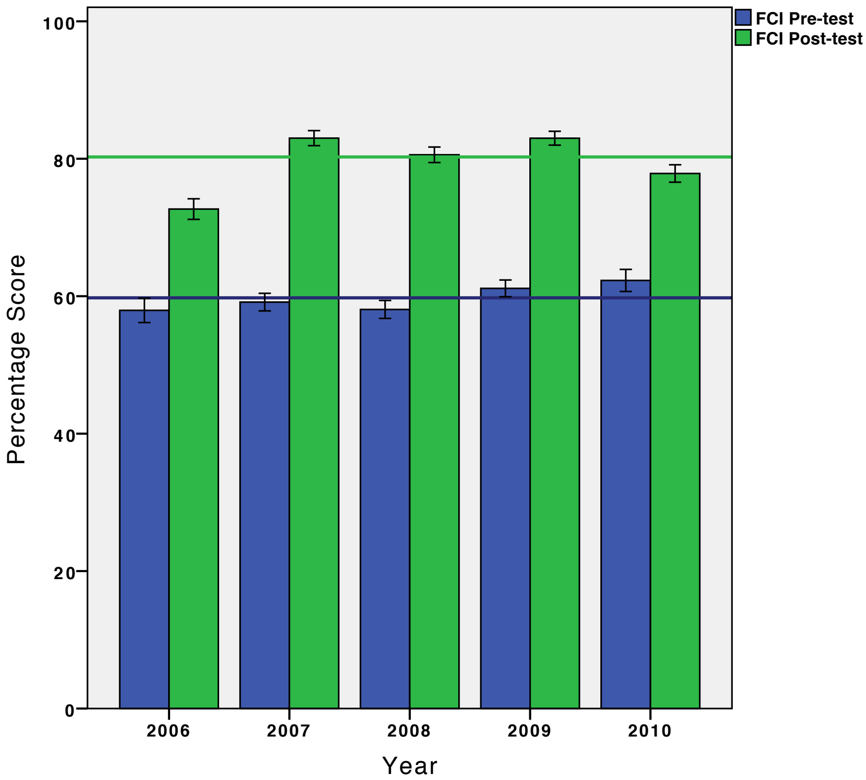
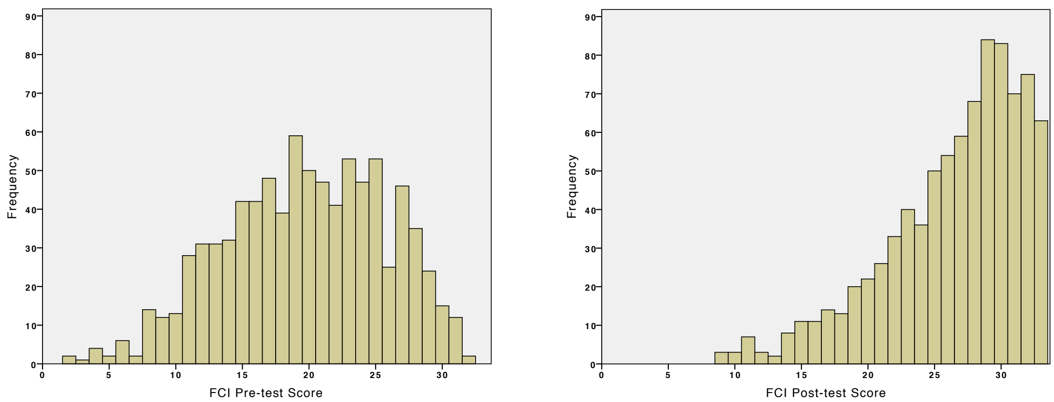


Figure 3.1: Mean pre-test and post-test scores for five consecutive years of the FCI_{ext} and the combined 2006-10 data set. Error bars represent the standard error on the mean. N(2006)=117, N(2007)=190, N(2008)=190, N(2009)=228, N(2010)=133. N=858 for combined data set of 2006-10. Horizontal lines indicate the average FCI pre-test (blue) and post-test (green) scores for the combined 2006-10 data set.

data collected from consecutive first year classes during this period. Average pre-test scores between successive cohorts showed no statistical differences at the 95% confidence level. The one exception to this was a statistically significant difference between the pre-test results for the 2008-09 and 2010-11 cohorts (p=0.032). Although there existed slight differences between post-test results, the five years between 2006 and 2010 were combined to create a larger sample size and higher confidence level in the statistical tests. The consistency of mean scores prior to instruction is evident in Figure 3.1, which shows the percentage pre-test and post-test scores for the five consecutive years of data alongside those for the combined data set. In total, 858 students completed both the pre- and post-tests to form the matched data set.

In each year, after one semester of teaching, students on average showed a statistically significant increase in their conceptual understanding. Looking at the combined data set, it was found that students' average pre-test score of 60(1)% increased to 80(1)% by the end of the course. Here, and in all subsequent values, the number in the brackets represents the standard error on the mean. It is worth noting that, on average, students are entering with a level of conceptual knowledge equal to that of the 'entry threshold' defined by Hestenes *et al* [46]. Each academic year showed large ranges in student scores, particularly in the pre-test, as shown in Figure 3.2 (a). The standard deviations of pre-test and post-test scores were 18.3% and 15.7% respectively. This increase in $\mathrm{FCI}_{\mathrm{ext}}$ scores offers reassurance about the course effect on student learning and is in line with previous studies using the FCI [37, 114].



(a) (b)

Figure 3.2: Histogram showing distribution of $\mathrm{FCI}_{\mathrm{ext}}$ (a) pre-test scores and (b) post-test scores for combined 2006-10 data set. N=858.

As a consequence of the decision made in 2011 to implement the original version of the FCI in place of the $\mathrm{FCI}_{\mathrm{ext}}$, results from the 2011-12 and 2012-13 cohorts were considered separately from previous years. In addition to this, from 2011 onwards, the course implemented an 'inverted classroom' methodology in lectures [63]. Looking first at the data sets independently, we can see that the students in both the 2011-12 and 2012-13 cohorts entered their degree programme with a level of conceptual understanding well above that of the 'entry threshold' for Newtonian mechanics (64(1)% and 69(1)% respectively) [46]. After one semester of teaching the students' scores exhibited a significant increase with the class average increasing to 84(1)% for 2011-12 and 85(1)% for 2012-13, comparable to the 'mastery threshold' of 85% [104]. Once again we saw a large range in student scores. The pre-instruction standard deviation of 19.7% was reduced to 14.3% post-instruction. Once again the two data sets were combined in order to create one larger data set (n=383) to increase the size and statistical confidence of the sample. Figure 3.3 shows pre-test and post-test results from the combined data set.

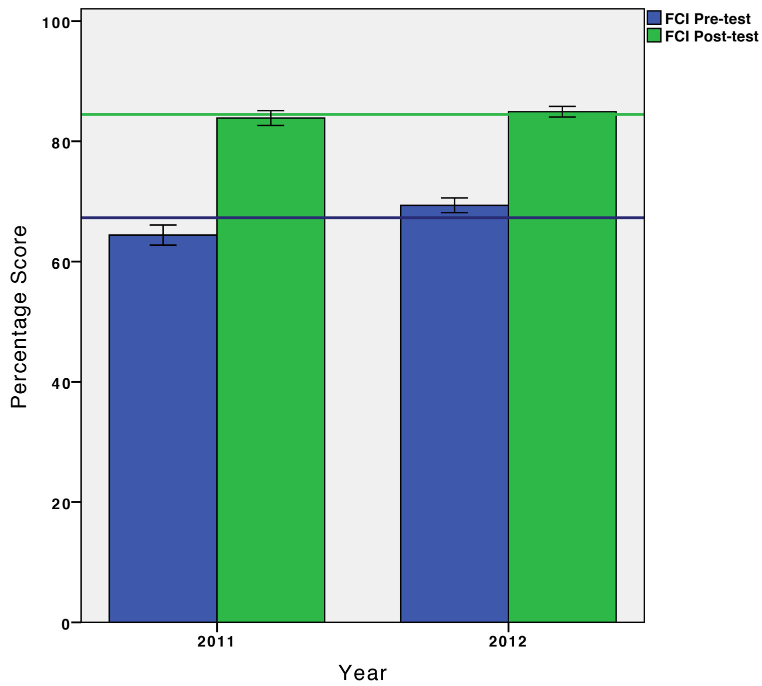


Figure 3.3: Mean pre-test and post-test FCI scores for 2011-12, 2012-13 and the combined 2011-13 data set. Error bars represent the standard error on the mean. N(2011-12)=161, N(2012-13)=222 and N(2011-13)=383 for pre- and post- instruction data. Horizontal lines indicate the average FCI pre-test (blue) and post-test (green) scores for the combined 2011-13 data set.

Over the course of the seven years in which the FCI or FCI_{ext} has been used to assess first year students, the average whole class pre-test scores have remained fairly consistent, with a slight increase in scores after the switch to the original version of the FCI. The mean pre-test scores for each year ranged from 57.9% to 69.4%. One feature of particular interest was the modal score on the post-instruction test. At the end of the ten week course we witnessed a ceiling effect. A ceiling effect on an assessment occurs when a high proportion of the population achieve maximum scores and there is very little variation amongst the top performing students. Data collected in the post-test showed that results were skewed towards the highest possible FCI scores. This post-test ceiling effect can be seen in Figure 3.2 (b). For 2006-10 the modal score was 88% and for 2011-13 the modal score was 97%.

The change in diagnostic test used, and therefore change in test questions, does not allow for a direct comparison of the seven years of data. One way to compare students' improved conceptual understanding among different year groups is to compare the average normalised gains of each population. Normalised gain is defined as the change in score from pre-test to post-test as a fraction of the total possible increase in score:

$$\langle g \rangle = \frac{\langle x \rangle_{\text{post}} - \langle x \rangle_{\text{pre}}}{100\% - \langle x \rangle_{\text{pre}}} \quad (3.1)$$

This is often considered a measure of instructional effectiveness, representing the fractional improvement in understanding, as described in Hake's study [44]. Presented in Table 3.1 are the average normalised gains for the first year cohort in each academic year. The average normalised gains of each population are particularly useful as students often begin with very different pre-test scores and a comparison of normalised gains allows us to compare the improvement in overall student FCI performance across year groups with different student populations.

Table 3.1: Average FCI normalised gains <g> for 2006-13 cohorts. The number in the brackets indicates the standard error on the mean.

Year	Normalised Gain
2006-07	0.33(4)
2007-08	0.58(2)
2008-09	0.54(2)
2009-10	0.54(2)
2010-11	0.38(3)
2011-12	0.55(3)
2012-13	0.44(3)

As discussed in Section 1.2.1, Hake's study found the average normalised gain for courses employing interactive engagement techniques to be <g>=0.48±0.14 compared to traditionally taught lecture courses which had a much lower than average normalised gain of <g>=0.23±0.04 [44]. Our results showed that all of the years considered demonstrated a normalised gain within the boundaries of a 'medium gain course' (0.7 ≥ <g> ≥ 0.3) as defined in Hake's study, suggesting an effective teaching/learning environment [44].

3.1.5 Gender differences in performance in 'Physics 1A'

In addition to looking at the scores of the cohort as a whole, a comparison was made between male and female students' FCI scores. Any differences seen can be denoted by the gender gap, G, where the difference between male and female mean scores is defined such that

$$G = <X_{male}> - <X_{female}>$$

where X represents either the pre- or post-test mean score. The convention adopted implies that a positive G indicates that male students are performing more highly than female students and vice versa.

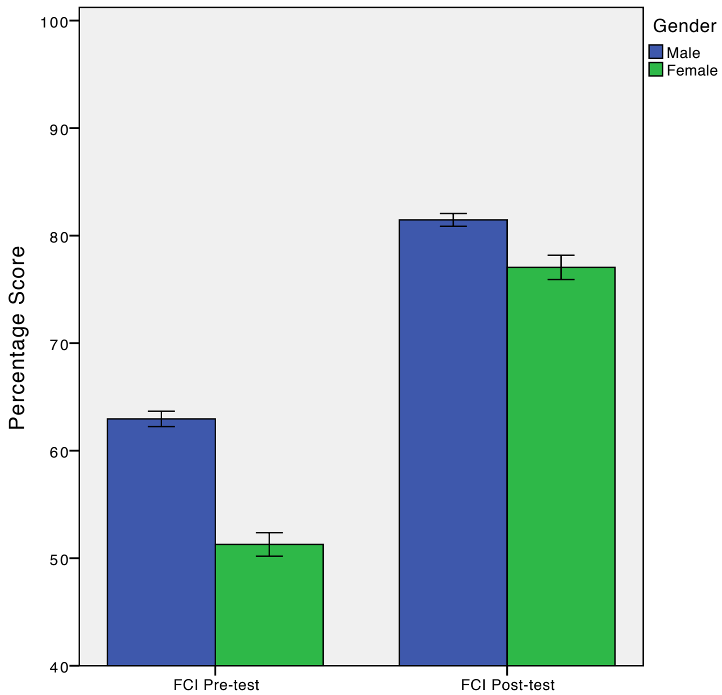
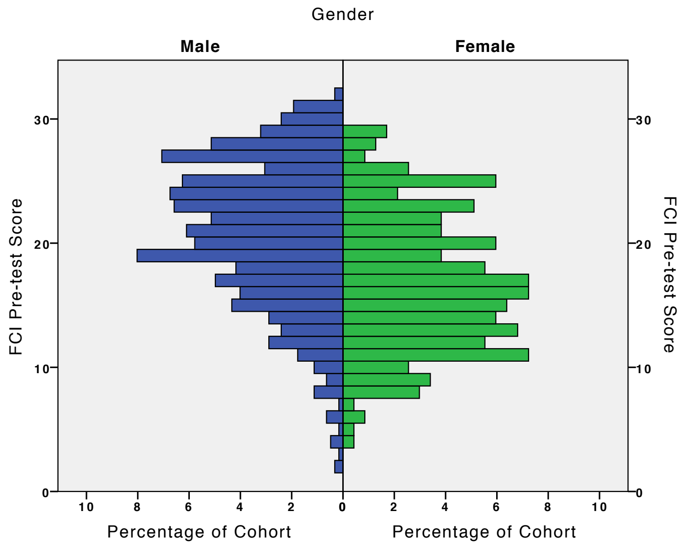


Figure 3.4: Mean pre-test and post–test FCI_{ext} scores for combined 2006-10 data set as a function of gender. Error bars represent the standard error on the mean. N(males)=623 and N(females)=235.

Results showed that a statistically significant gender gap existed between first year undergraduate physics students as measured by both versions of the FCI. As can be seen from Figure 3.4, which shows results from the combined 2006-10 data set, the males outperformed the females, scoring on average 12% higher than incoming females. This gender difference was found to be statistically significant using a Mann-Whitney U test (p<0.001). This suggests that females may enter university with a significantly lower understanding of key Newtonian concepts. After one semester of teaching, both cohorts improved greatly, as expected by the overall improvement of the whole class cohort. However, male students still performed on average 4% higher than females and the gender gap remained statistically significant (p=0.002).

(a)



(b)

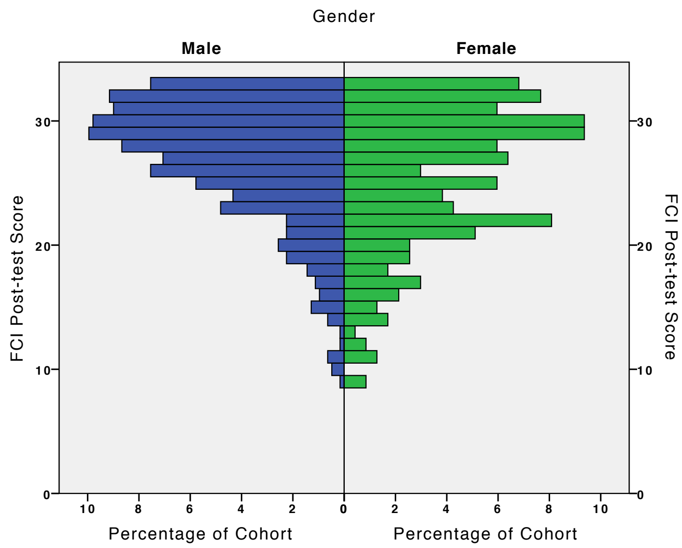


Figure 3.5: Gender distribution of FCI_{ext} (a) pre-test scores and (b) post-test scores for combined 2006-10 data set. N(males)=623 and N(females)=235.

Further investigation of the pre-test scores of each gender cohort showed a mixed familiarity with Newtonian concepts, with both genders having a wide distribution of test scores, as illustrated in Figure 3.5 (a). Although it decreased greatly, the significant gender gap persisted in the post-test results. Once again, the post-test score distribution seen in Figure 3.5 (b), showed a slight ceiling effect for both cohorts.

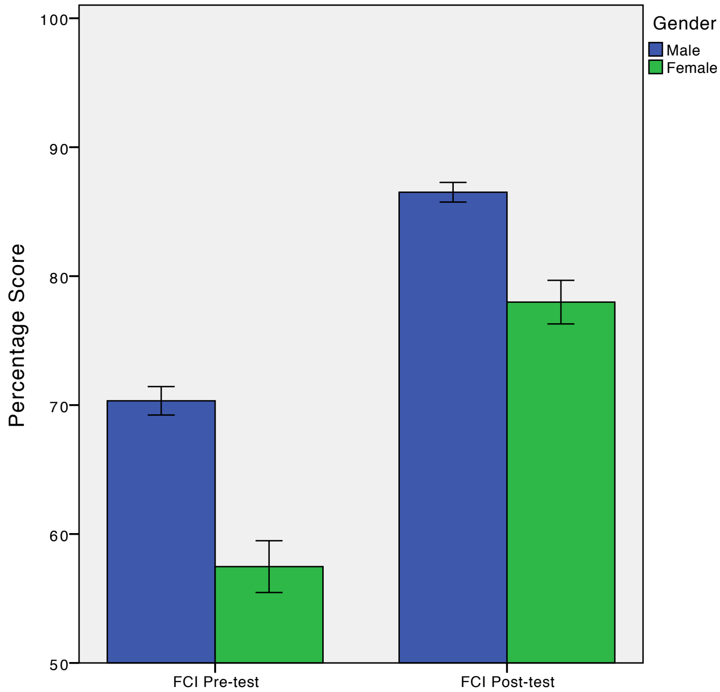


Figure 3.6: Mean pre-test and post-test FCI scores for combined 2011-13 data set as a function of gender. Error bars represent the standard error on the mean. N(males)=292 and N(females)=91

Looking at data from 2011-13, shown in Figure 3.6, we see a similar pattern in performance with respect to gender. There remained a persistent gender gap at the end of the semester. Males outperformed female students by 11% as measured in week one, reducing to just under 8% by the end of the semester, but remaining statistically significant (p=0.012). This result is in partial agreement with the study by Docktor and Heller who also found that the gender gap remained significant post instruction, but the gender performance gap measured in their study did not show as large a decrease between the pre- and post-tests (15.3% to 13.4%) [62]. Mazur *et al* found a statistically significant gender gap of comparable magnitude (9-15%) when students started the course [55]. However, in contrast to our finding, after one semester of teaching they witnessed a complete elimination of the FCI gender performance gap.

A point of interest is the comparison of normalised gains of both genders. In most years, with the exceptions of 2006 and 2011, females had an equal or higher normalised gain than males, as shown in Table 3.2. Over the seven years studied, females showed an average normalised gain of 0.52 compared to 0.49 for the males. This higher normalised gain may go towards explaining the decrease in the gender gap over the semester, with females having a greater improvement in their conceptual understanding. Once again all cohorts exhibited high gains consistent with successful interactive courses.

Table 3.2: Average FCI normalised gains <g> for male and female students between 2006-13. The numbers in the brackets indicate the standard error on the mean.

Year	Normalised Gain Male Students	Normalised Gain Female Students
2006-07	0.33(5)	0.32(5)
2007-08	0.56(3)	0.64(4)
2008-09	0.54(3)	0.54(4)
2009-10	0.52(3)	0.62(4)
2010-11	0.36(4)	0.42(4)
2011-12	0.57(4)	0.52(6)
2012-13	0.44(4)	0.47(5)

3.1.6 Changes in the gender gap over time

As well as measurements of the gender performance of each academic year group, the progression of the gender gap over the course of a semester was studied. By comparing the FCI pre- and post-test scores for each gender for the matched data sets, an indication of the change in the gender gap over a semester could be determined. The change in the gender gap was defined as

$$\Delta \text{ G} = \text{G}_{(post-test)} \text{ - } \text{G}_{(pre-test)}$$

where a positive value of Δ G indicates a that gender gap between male and female students has widened, and a negative value indicates that the gender gap has narrowed.

For all years in question the change in the gender gap was negative (Figure 3.7), indicating that the magnitude of the gender gap decreased after one semester of teaching. The average change in the gender gap was calculated to be -6.38%. The shift in the gender gap for each year lay within one standard deviation of the average change in the gender gap. This further allowed for 2006-10 data to be combined into a single data set covering the whole cohort. This observed change in the difference between male and female performance is contrary to conclusions previously found by Docktor and Heller [62]. Over a decade of teaching they found that, while some semesters showed a decreasing gender gap, others resulted in a widening of the gender gap.

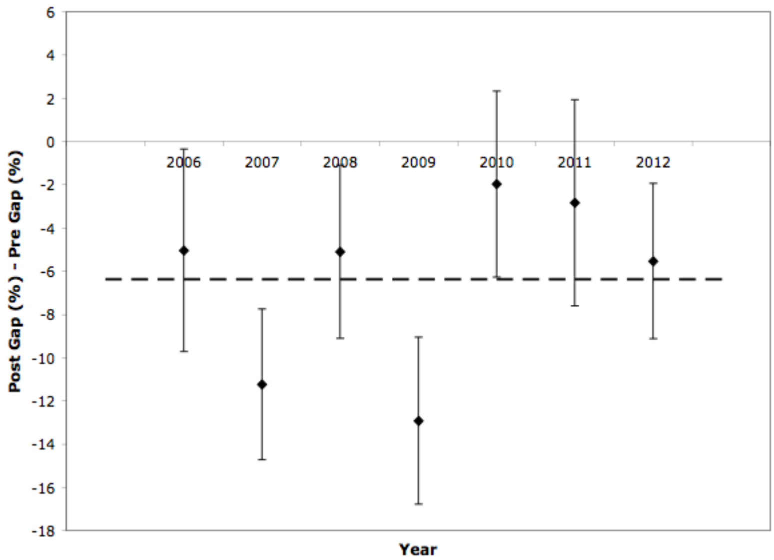


Figure 3.7: Percentage change in the FCI gender gap for students enrolled on 'Physics 1A' between 2006-13. Error bars represent the standard error on the mean. Horizontal line indicates the mean percentage gender gap over the seven years analysed.

3.1.7 Comparison of physics majors and non-majors

As mentioned earlier, only around 50% of students who take the 'Physics 1A' course are enrolled on a degree programme where their intention is to study for a physics degree. Throughout all following sections, these students are referred to as physics 'majors'. The term 'non-majors' is used to refer to students who are not intending to complete a physics degree, but are instead choosing to enrol on the 'Physics 1A' course as an outside course. Of the 858 students who completed both the pre- and post-test FCI_{ext} between 2006-10, 854 could be classified as 'major' or 'non-major'; 432 majors and 422 non-majors. The remaining four students not included in the final analysis were visiting students on exchange from international universities.

Figure 3.8 shows the pre-test and post-test percentage scores for majors and non-majors averaged over all years between 2006 and 2010. Several points of interest can be seen from this figure. First, a direct comparison of the two populations suggests that physics majors and non-majors entered the course with different initial levels of conceptual understanding of force and motion. Results of a Mann-Whitney U test indicated a statistically significant difference ($p<0.001$) between majors and non-majors in the pre-test FCI_{ext}. When students completed the test in week one of undergraduate teaching, majors on average scored 62(2)%, compared to non-majors who scored on average 57(2)%. Majors performed on average 5% higher than non-major students. Another interesting point is the change in the performance gap between major and non-major students. This difference decreased to 2% after one semester of teaching, suggesting that non-majors had a higher percentage improvement over the course. Despite this, the gap between the two cohorts remained marginally significant ($p=0.030$).

The 2011-13 cohorts demonstrated a similar pattern in physics majors' and non-majors' scores. A total of 383 students could be classified as 'major' or 'non-major'; 159 majors and 224 non-majors. The combined 2011-13 cohort contained a lower proportion of physics majors than the previous combined data set, with 42% of students enrolling on a physics degree programme. Results from 2011-13 majors and non-majors, illustrated in Figure 3.9, showed higher mean pre- and post-instruction scores compared to 2006-10. However, a different version of the test had been used in these two year groups and new teaching methods introduced, so it is not valid to draw any statistical comparison between these cohorts. When comparing majors and non-majors in 2011-13, a statistically significant difference was witnessed between the two groups prior to instruction ($p=0.042$). Students who declared themselves as physics majors at the start of the semester scored on average 4.5% higher on the FCI than those who completed the course as an elective. Major students scored 70(2)% on the pre-test compared to non-majors who scored 65(1)%. This statistically significant gap was completely closed by the end of the semester and was no longer significant ($p=0.456$). Both major

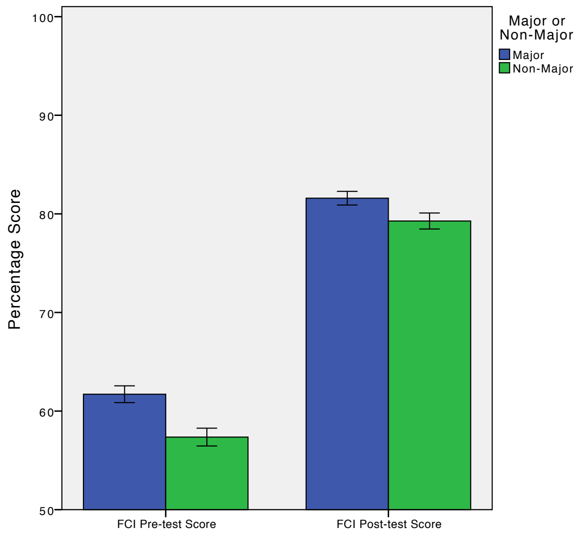


Figure 3.8: Mean pre-test and post-test FCI_{ext} scores for majors and non-major in 2006-10 combined data set. Error bars represent the standard error on the mean. N(majors)=432 and N(non-majors)=422.

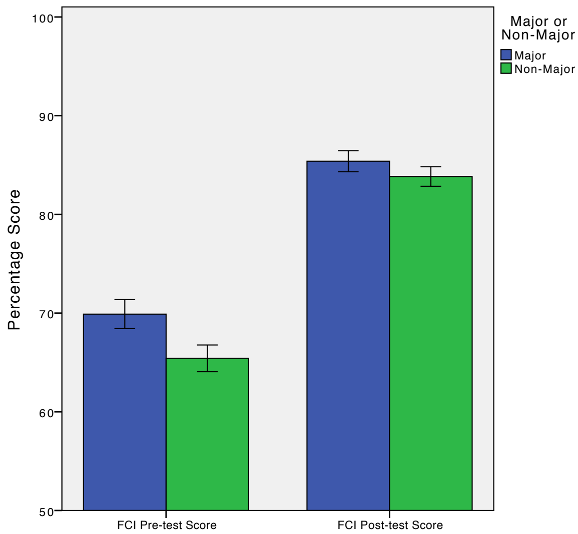


Figure 3.9: Mean pre-test and post-test FCI scores for majors and non-majors in 2011-13 combined data set. Error bars represent the standard error on the mean. N(majors)=159 and N(non-majors)=224.

and non-major students showed improved post-test scores of 85(1)% and 84(1)% respectively. We can note that the majors and non-majors had now reached a level of Newtonian 'mastery'.

Although this pre-test difference between the two groups may not be entirely unexpected, it is interesting to note that major and non-major students enter the course having met the same minimum entrance qualifications and it is simply due to their personal choice that non-majors have chosen not to enrol on a physics degree but instead pursue another discipline. The higher pre-test scores between 2011-13 is a reflection of the higher performance level of the whole cohort on the original FCI compared to the FCI_{ext}. Irrespective of this change, both combined data sets showed comparable percentage differences at the start of the semester.

The normalised gains of each cohort are shown in Table 3.3. There was a large variation in normalised gain amongst individual years between 2006-13. In almost all cases cohorts of students not enrolled on the physics degree programme had a higher normalised gain than physics majors. The one exception was 2009-10 where physics majors had a significantly higher normalised gain than physics non-majors.

Table 3.3: Average FCI normalised gains <g> for major and non-major students between 2006-13. The numbers in the brackets indicate the standard error on the mean.

Year	Normalised Gain Majors	Normalised Gain Non-Majors
2006-07	0.36(6)	0.39(4)
2007-08	0.57(4)	0.59(3)
2008-09	0.51(3)	0.61(3)
2009-10	0.59(3)	0.50(4)
2010-11	0.33(6)	0.40(4)
2011-12	0.52(6)	0.58(4)
2012-13	0.43(5)	0.46(4)
Combined 2006-2010	0.50(4)	0.51(4)
Combined 2011-2013	0.47(4)	0.51(3)

As shown earlier in this chapter, the 2011-12 cohort had a much higher normalised gain than that of 2012-13 students. Majors in 2011-12 had a normalised gain of 0.52(6) compared to non-majors who had a normalised gain of 0.58(4). This difference was not significant at the 95% level. Similarly, in 2012-13, the majors had a lower normalised gain of 0.43(5) compared to non-majors who had a gain of 0.46(4). Results from the 2011-13 combined data set showed a slightly higher, but not significant, normalised gain for non-major students (<g>=0.51(3) for non-majors and <g>=0.47(4) for majors). This is in agreement with the fact that we see a closure of the significant difference between major and non-major students' performance on the post-instruction FCI test.

3.1.8 Gender analysis of physics majors and non-majors

After observing this difference between major and non-major students prior to instruction, further investigations were undertaken to determine whether there was a gender dimension to these results. Is the gender difference a result of a difference between the performance male and female non-major students? Do female majors and non-majors have a significant difference in levels of conceptual understanding of Newtonian mechanics? The matched data were coded for both subject major and gender. Once more, data for the five years 2006-10 were combined to form one data set and considered separately from data from 2011-13. Table 3.4 shows the number of students completing the FCI in each cohort for each academic year.

Table 3.4: Number of students split by subject major and gender who completed both pre-test and post-test FCI assessments.

Year	Male Majors	Male Non-Majors	Female Majors	Female Non-Majors
2006-07	41	37	22	16
2007-08	54	83	25	28
2008-09	84	55	26	22
2009-10	98	78	24	27
2010-11	41	47	17	28
2011-12	50	66	18	27
2012-13	72	104	19	27
Combined 2006-2010	318	300	114	122
Combined 2011-2013	122	170	37	54

Examination of the pre-test and post-test scores for the FCI depicted in Figure 3.10 illustrated that there was a marked difference in performance depending on both subject major and gender. Between 2006-10 male majors significantly outperformed all other cohorts both pre- and post-instruction. The significant difference that existed between male majors and male non-majors in the pre-test ($p<0.001$) remained in the post-test results ($p=0.023$), although the gap decreased. Similarly, a statistically significant gap ($p<0.001$) existed between male majors and female majors both at the start of the semester (15.2%) and following a semester of teaching (5.14%). A statistical difference with p value of 0.002 was found post-instruction. Unlike for males, female majors and female non-majors showed no statistically significant differences in either pre-test (p=0.877) or post-test scores (p=0.707), with scores almost identical for both tests (51(2)%). Incoming male non-majors significantly outperformed female non-majors by approximately 10% (p<0.001). After a semester of teaching, the gap narrowed (3.3%) but remained significant with a p value of 0.047.

Looking at the 2011-13 combined data shown in Figure 3.11, it can be seen that male majors outperformed all other sub-cohorts both prior to and post instruction. This is in agreement with the results seen in the previous five years. However, the difference between male majors and male non-majors was not statistically significant. The significant difference (p=0.001) that existed between male majors, who scored 73(2)% at the beginning of the semester, and female majors, who scored 61(3)% pre-test, decreased but remained significant (p=0.018), even after ten weeks of teaching. A similar pattern was seen when comparing male and female non-major students. Male non-majors performed significantly higher (p=0.028) than female non-majors pre-test, 69(1)% and 55(3)% respectively. This significant gap persisted in the post-instruction assessment but reduced to 9.5%. The difference between male majors and female non-majors was very evident. Male majors scored 17.7% higher than female non-majors who under-performed compared to all other subgroups. Despite this, there was no significant difference between female majors and female non-majors either pre-test or post-test.

When data for each year's cohort were examined separately, large variations were seen from year to year, particularly with the performance of female non-majors compared to other subgroups (Figures F.1 - F.7 in Appendix). In four of the seven years examined, female non-majors underperformed compared to all other cohorts in the pre-test administration of the FCI. The three years in which this was not the case were 2006-07, 2007-08 and 2010-11. In 2010-11 female non-majors outperformed both male non-majors and female-majors and there was no statistical significance between any groups in the FCI pre-test, nor was there any difference between male non-majors, female majors or female non-majors at the end of the semester. These fluctuations could be a consequence of the

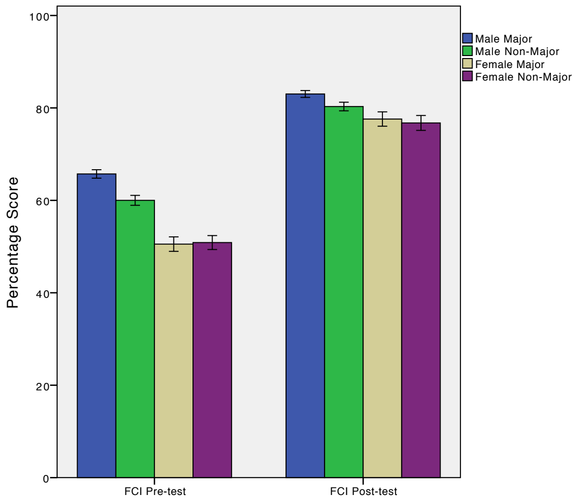


Figure 3.10: Mean pre-test and post-test FCI_{ext} scores for male and female majors and non-majors in combined 2006-10 data set. Error bars represent the standard error on the mean.

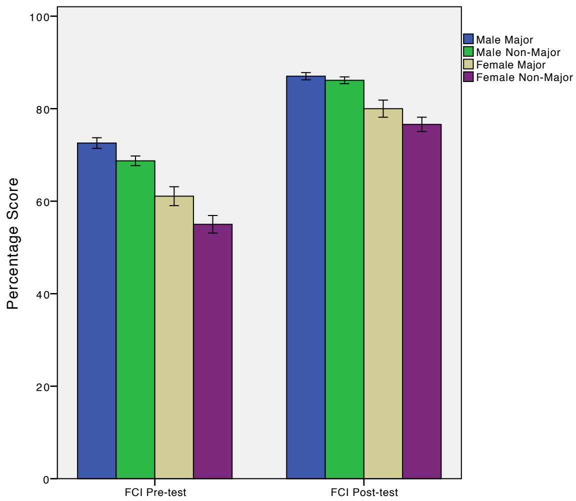


Figure 3.11: Mean pre-test and post-test FCI scores for male and female majors and non-majors in combined 2011-13 data set. Error bars represent the standard error on the mean.

variation in number of female students who participated in the test. In 2010-11, the proportion of female students rose to almost 30% compared to 20-25% seen in previous years. Of these females, almost 60% were non-major.

Table 3.5 shows the normalised gains for each cohort. As seen when comparing male and female students' conceptual understanding performance in section 3.1.4, female majors had a consistently higher normalised gain after one semester of teaching than male majors. This difference was statistically significant (p=0.019) in 2009-10 but not in other years.

Table 3.5: FCI normalised gains <g> for each cohort split by subject major and gender. The numbers in brackets indicate the standard error on the mean.

Year	Male Majors	Male Non-Majors	Female Majors	Female Non-Majors
2006-07	0.36(8)	0.41(5)	0.37(7)	0.35(5)
2007-08	0.52(6)	0.56(3)	0.66(5)	0.65(5)
2008-09	0.50(4)	0.61(4)	0.53(6)	0.61(6)
2009-10	0.55(4)	0.49(5)	0.73(5)	0.53(7)
2010-11	0.31(7)	0.39(5)	0.39(6)	0.41(5)
2011-12	0.51(7)	0.61(3)	0.55(8)	0.49(9)
2012-13	0.40(6)	0.47(5)	0.54(7)	0.42(6)
Combined 2006-2010	0.48(2)	0.51(2)	0.55(3)	0.52(3)
Combined 2011-2013	0.44(5)	0.53(4)	0.55(5)	0.46(5)

3.1.9 Relationship between pre-test and post-test scores

Having found evidence of gender disparity in average scores on a conceptual understanding test of Newtonian mechanics which remained statistically significant even after a semester of teaching, it was investigated whether male and female students with similar pre-test scores also completed the course with equivalent post-test scores. Results discussed in section 3.1.5 have shown both cohorts to have high normalised gains between the two FCI assessments, with females having equal or higher normalised gains in all but two years (2006-07 and 2011-12). In this section an attempt was made to understand if this higher normalised gain was the result of a high level of improvement by female students in the lower quartiles of the class. Similarly, is the gender difference more distinct in a particular quartile of students?

In order to understand the narrowing of the gender gap further, students were binned into four bins, hereafter referred to as quartiles, based on their pre-test FCI scores. Each quartile contained an approximately equal population size. For each performance quartile the subsequent average post-test score was calculated separately for both the male and female populations in each bin. Variations in size of quartile populations resulted from large numbers of students scoring on the boundary levels of each quartile.

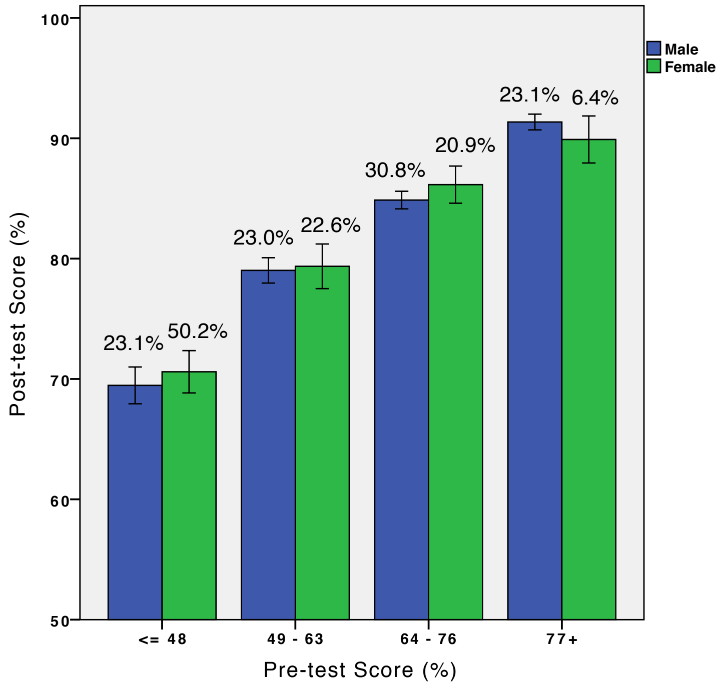


Figure 3.12: FCI_{ext} post-test scores as a function of male and female pre-test performance quartiles for combined 2006-10 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=623 and N(females)=235.

It can be seen from Figure 3.12, which shows 2006-10 male and female 'Physics 1A' students' average post-test score as a function of their pre-test quartile, that students who obtained comparable pre-test scores also had similar post-test scores. Analysis indicated that there were no statistically significant differences between genders in any of the four quartiles. Comparisons of this data with the entry threshold value for Newtonian mechanics set out by Hestenes *et al* (60%), showed that on average all subgroups of students achieved this level by the end of the semester of teaching [46]. In fact, the mean scores of the top two quartiles exceeded that of the Newtonian 'mastery' benchmark (85%) [104].

Data combined from the two years 2011-12 and 2012-13 (Figure 3.13) showed that, although there was no evident gender gap between males and females in the top three quartiles of the class, female students in the lowest quartile significantly underperformed in the post-test compared to their male counterparts (p=0.001). Females in this quartile scored an average of 68(1)% post-test compared to males who scored 75(1)%. If we then consider all of the data sets, we see that females in the lowest quartile scored lower than males in five of the seven years analysed (Figures G.1 - G.7).

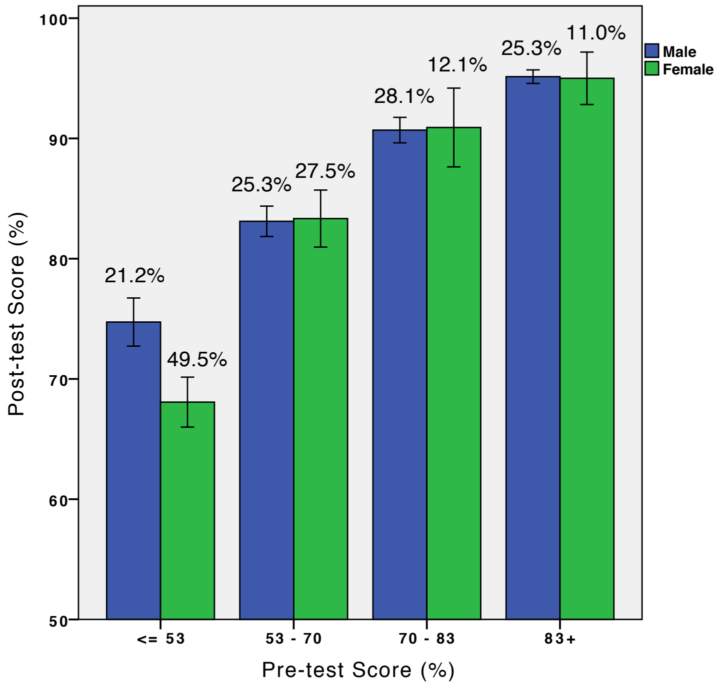


Figure 3.13: FCI post-test scores as a function of male and female pre-test performance quartiles for combined 2011-13 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=292 and N(females)=91.

One of the key features of this analysis is the distribution of the male and female cohorts across the four performance quartiles, as defined by pre-test scores. A comparison of the two genders between 2006-10 showed that a much higher percentage of female students fell into the lower pre-test score quartiles, with 50% of females scoring less than 48% on the FCI when they began their undergraduate degree, compared to only 23% of males (Figure 3.12). Conversely, a higher proportion of male students (54%) achieved FCI pre-test scores in the highest two quartiles compared to only 27% of females. In each year the male population is distributed relatively evenly across the four pre-test quartiles. The proportion of the female students in each quartile bin decreased as we moved up the quartiles. This was also the case between 2011-13 as seen in Figure 3.13. Once again 50% of the female population performed in the lowest performance quartile compared to 21% of male students.

Students were rebinned into quartiles of approximately equal population size based on their post-test FCI scores, using the same approach discussed earlier in this section. If we look at the 'churn' in the distribution of students between pre- and post-testing, shown in Tables 3.6 and 3.7, we find that the majority of students who scored in the bottom quartile pre-test also scored in the bottom quartile at the end of the semester. In fact, the percentage of females who remained in the lowest quartile is higher than that of males, particularly in the combined 2011-13 data set (Tables 3.8 and 3.9).

These results may suggest that the initial gender gap seen between incoming students is a result of this high proportion of lower scoring females, and that the reduction in the gender gap witnessed after one semester of teaching may be attributed to the fact that female students have a higher normalised gain than males. Nevertheless, a high percentage of these females remain in the lowest quartile. It may be tempting to suggest that this is due to these students simply being weaker overall at the point of entry. Looking at students' educational backgrounds and school leaving qualifications, we find little evidence for this. Although it is difficult to obtain data informing us of the physics and mathematics modules completed by students in their final years of secondary school, entrants to the physics course arrive with very similar, high level qualifications. In recent years students are applying with straight As in their final school examinations. Despite this, comparing students with the same grades may not accurately reflect their prior exposure to concepts of Newtonian mechanics. In fact evidence suggests that, across the UK, females outperform male students in school-leaving examinations, including physics [115, 116].

Table 3.6: FCI_{ext} quartile distribution of males in combined 2006-10 cohort. N(males)=623. The percentage values represent the proportion of male students from each pre-test quartile present in each post-test quartile.

	FCI Post-test	Q4 (≤23)	Q3 (24-27)	Q2(28-30)	Q1(31+)
FCI Pre-test		Nm=132	Nm=154	Nm=177	Nm=160
Q4 (≤16)	N=144	71 **49.3%**	33 22.9%	27 18.8%	13 9.0%
Q3 (17-20)	N=143	40 33.8%	44 **43.2%**	37 17.6%	22 5.4%
Q2 (21-25)	N=192	18 9.4%	60 31.3%	65 **33.9%**	49 25.5%
Q1 (26+)	N=144	3 2.1%	17 11.8%	48 33.3%	76 **52.8%**

Table 3.7: FCI_{ext} quartile distribution of females in combined 2006-10 cohort. N(female)=235. The percentage values represent the proportion of female students from each pre-test quartile present in each post-test quartile.

	FCI Post-test	Q4 (≤23)	Q3 (24-27)	Q2(28-30)	Q1(31+)
FCI Pre-test		Nm=84	Nm=45	Nm=58	Nm=48
Q4 (≤16)	N=118	61 **51.7%**	18 15.3%	24 20.3%	15 12.7%
Q3 (17-20)	N=53	15 28.3%	16 **30.2%**	14 26.4%	8 15.1%
Q2 (21-25)	N=49	7 14.3%	10 20.4%	14 **28.6%**	18 36.7%
Q1 (26+)	N=15	1 6.7%	1 6.7%	6 40.0%	7 **46.7%**

Table 3.8: FCI quartile distribution of males in combined 2011-13 cohort. N(males) =292. The percentage values represent the proportion of male students from each pre-test quartile present in each post-test quartile.

	FCI Post-test	Q4 (≤23)	Q3 (24-27)	Q2(28-29)	Q1(30)
FCI Pre-test		Nm=61	Nm=106	Nm=88	Nm=37
Q4 (≤16)	N=62	34 **54.8%**	19 30.6%	5 8.1%	4 6.5%
Q3 (17-21)	N=74	25 33.8%	32 **43.2%**	13 17.6%	4 5.4%
Q2 (22-25)	N=82	2 2.4%	38 46.3%	34 **41.5%**	8 9.8%
Q1 (26+)	N=74	0 0.0%	17 23.0%	36 48.6%	21 **28.4%**

Table 3.9: FCI quartile distribution of females in combined 2011-13 cohort. N(female)=91. The percentage values represent the proportion of female students from each pre-test quartile present in each post-test quartile.

	FCI Post-test	Q4 (≤23)	Q3 (24-27)	Q2(28-29)	Q1(30)
FCI Pre-test		Nm=45	Nm=21	Nm=17	Nm=8
Q4 (≤16)	N=45	35 **77.8%**	8 17.8%	2 4.4%	0 0.0%
Q3 (17-21)	N=25	8 32.0%	10 **40.0%**	5 20.0%	2 8.0%
Q2 (22-25)	N=11	1 9.1%	3 27.3%	4 **36.4%**	3 27.3%
Q1 (26+)	N=10	1 10.0%	0 0.0%	6 60.0%	3 **30.0%**

3.1.10 Effect of reordering questions

An alternative question asked of the observed FCI results is whether the gender gap between male and female first year students is a 'real' gender feature or is it an inherent feature of the concept inventory? Gray *et al* found that the order of questions in a conceptual test could have a statistically significant effect on students' performance [117]. Their study investigated the effect of reordering two FCI questions relating to Newton's Third Law. The first of the two Newton's Third Law questions considered the forces between two accelerating objects in contact whilst the second question referred to the same two objects moving at a constant speed in contact with each other. When presented with the accelerating scenario first, almost 10% of respondents stated that in the second scenario there existed no force between the two objects travelling at a constant speed. In contrast, when presented with the constant speed scenario prior to the accelerating question, no students indicated that they believed there to be no interactive forces between the objects. The statistically significant difference found between these results suggests that question order is an important issue for consideration. In order to explore this, two versions of the FCI were used in the 2011-12 and 2012-13 academic years to test whether question order had an effect on students' attainment. One of these versions was the original FCI. The second version featured the original FCI with the questions reordered randomly (but clusters of questions that are connected to one another or refer to the same diagram were maintained). For the full order of FCI questions in the random order test see Appendix A. Students were randomly assigned to either the 'Original order' or 'Random order' FCI and took the same version of the test both pre- and post- instruction to ensure a fair comparison of any effects of this intervention.

In 2011-12 and 2012-13 the male students completing the 'Original order' FCI significantly outperformed female students completing the same test in both the pre-test and post-test results (Table 3.10). There was no such significant difference if we compared genders in the 'Random order' FCI for 2011-12, although we saw a gender difference in 2012-13 students. Interestingly, a difference in performance was found when comparing males students across the two test versions in 2011-12. The male students who completed the 'Random order' FCI performed significantly lower than the 'Original order' male students (p=0.020) in 2011-12. This difference was eliminated by the end of the semester,

Table 3.10: FCI percentage pre-test and post-test scores for 'Original order' and 'Random order' FCI as a function of gender. Numbers in brackets represent the standard error on the mean. In 2011-12 N(males 'Original order')=62, N(males 'Random order')=54, N(females 'Original order')=21 and N(females 'Random order')=24. In 2012-13 N(males 'Original order')=90, N(males 'Random order')=86, N(females 'Original order')=22 and N(females 'Random order')=24.

			Original Order		Random Order	
Year	Gender	N	Pre-test (%)	Post-test (%)	Pre-test (%)	Post-test (%)
2011-12	Male	116	71(3)	87(2)	63(3)	85(2)
	Female	45	54(5)	76(4)	59(4)	80(3)
2012-13	Male	176	72(2)	87(1)	73(2)	87(1)
	Female	46	56(1)	78(3)	60(3)	78(3)

with both groups scoring above 85%. Conversely, females students scored slightly higher in the 'Random order' test in the same year, but the difference was not significant at the 95% level. When this study was repeated in the 2012-13 academic year, no statistical differences were seen in student performance across the two versions of the FCI. This may imply that the results seen in the previous year were a result of differences in the cohorts who had been randomly grouped into the two tests. Whether this change in male performance is the result of a change in the cohort is uncertain. It is possible that it is a reflection of changes in the make up of the physics class compared to previous years; the 2012-13 class had a noticeably higher proportion of male students compared to previous years (Table 2.1). Whilst the absolute number of female students in the class remained almost constant over the two years, the number of male students increased greatly. Additionally, the 2012-13 cohort had an increase in the number of non-major students enrolling in 'Physics 1A'. Although differences can be seen between subgroups across the two tests, this does not necessarily imply causation and may simply be the result of a statistical fluctuation.

3.2 A three institution comparison of gender differences in conceptual understanding

3.2.1 Aim of quantitative study

In addition to exploring gender differences in students entering first year at the University of Edinburgh, it is important to understand whether these apparent differences in male and female conceptual understanding are reflected in results nationwide. In this section results are presented from a comparison of the gender differences witnessed in performance in Newtonian mechanics at three UK universities in 2011-12[1]. Here results are detailed from pre- and post-instruction testing using the FCI at three different UK institutions: University of Edinburgh, University of Hull and University of Manchester [118]. This study aimed to not only evaluate the existence, and possible persistence, of a performance gender gap in introductory physics courses, but also to look at the test questions individually for any significant gender differences that are common across all three institutions.

3.2.2 Implementation of FCI

All three universities have been using the FCI (or the variant FCI_{ext}) as an assessment instrument within their first year introductory physics courses for a number of years. As mentioned earlier in this chapter, the University of Edinburgh had been using the FCI_{ext} prior to 2011. All three institutions aligned their processes in the 2011-12 academic year, on which the results in this section are based, with each group using the original version of the FCI with first year students.

There existed slight variations in the implementation of the FCI at each university, as detailed in Table 3.11, for reasons of practicality and course assessment. Both Hull and Manchester universities administered the FCI in paper form and gave no course credit for student participation, with Manchester also enforcing a time limit of sixty minutes. As mentioned in the previous section, students at the University of Edinburgh completed the inventory test online with

[1]This study was undertaken in conjunction with the University of Hull and the University of Manchester. Data from first year undergraduate students at the three universities were compiled by, and the statistical analysis conducted by the thesis author.

a time limit of ninety minutes. The best of their two attempts contributed an overall 3% to their final mark.

Table 3.11: FCI implementation details at participating universities and contribution of assessment to final course mark.

University	FCI used since	Delivery mechanism	Time limit	Timing pre- / post-	Contribution to final mark (%)
Edinburgh	2006[a]	Online	90 mins	weeks 1 / 8	3
Hull	2008[b]	Paper	none	weeks 0 /10	0
Manchester	2008	Paper	60 mins	weeks 0 / 6	0

[a] Between 2006 and 2010 a variant of the FCI was used with additional questions (FCI_{ext}).
[b] Matched pre- and post-data were collected for the first time in 2011.

3.2.3 Institutional contexts

All three universities require students to have fulfilled specific school qualifications in physics and mathematics before enrolling on first year undergraduate physics courses, but there were slight differences between entry qualifications and cohorts amongst the universities. Students in this study have completed a range of types of school leaving qualifications and therefore their prior knowledge of the subjects could vary depending on their examination modules.

Edinburgh

As outlined in section 2.1.1, the School of Physics and Astronomy at the University of Edinburgh has an average intake of approximately 120 students enrolling on the physics degree programme, with a total class size of 200-300 students in the first semester physics course ('Physics 1A'). Of these students, approximately 24% are female. The Scottish Bachelor's degree extends to four years, with the first year presenting a broader curriculum than in England. In this first year a third of students' courses are filled by those of the students' choosing, with the remaining two thirds comprising compulsory courses for their degree programme. In 2011-12 academic year the first semester course included PeerWise and an 'inverted' classroom approach in lectures [63, 64, 95, 96].

Hull

The first year physics course at the University of Hull has a general intake of approximately 70 students, most of whom enrol with A-level qualifications. Of the students in 2011-12, 10% were female. The introductory physics course runs for 10 weeks in the first semester and employs both formal instruction and interactive engagement teaching methodologies. The formal instruction is based on a structured approach to the use of multiple representations in constructing models, with the role of representations in evaluating, describing, analysing and solving problems being emphasised. Results from the implementation of this teaching model are discussed further in a study by Sands and Marchant [119].

Manchester

At the University of Manchester between 230 and 290 students enrol on the first year undergraduate programme each year. Of these students approximately 20% are female. The 11-week Newtonian mechanics based course taken by students in their first year employs interactive electronic voting, Peer Instruction and Just-in-Time Teaching [96]. The implementation of these teaching techniques has shown an increase in both student examination performance and overall course satisfaction [120, 121].

3.2.4 Quantitative results from the FCI at three institutions

The FCI was administered to each cohort before and after relevant instruction, during which time all content tested by the FCI was covered. In results presented in the following section, only matched pairs of data (data for students who had taken both the pre- and post-instruction tests) were included. Once again, this led to the sample size stated for each institution being lower than the total class size. For the members of each cohort undertaking both a pre- and post-instruction test, the pre- and post-instruction average percentage scores were calculated.

The three populations started the course with very different average pre-test scores; ranging from 59(3)% to 76(1)%. A comparison between the cohorts was made by calculating the overall learning gains for each university. Cohorts from all three institutions showed substantial normalised gains (<g>) on the FCI, ranging from 0.41 to 0.55, comparable with those seen on ‘reformed’ courses in studies reported in the literature [44]. These gains, presented in Table 3.12, provided evidence to suggest that interactive methodologies used in these lecture courses have led to high levels of student performance.

Table 3.12: Three institution student performance on the pre-test and post-test FCI in 2011-12. Values in brackets are the standard error on the mean.

Institution	Group	N		$\langle x \rangle (\%)$	$\langle g \rangle$	G	p	ΔG
Edinburgh	Whole class	161	Pre	64(2)				
			Post	84(1)	0.55			
Hull	Whole class	46	Pre	59(3)				
			Post	76(2)	0.41			
Manchester	Whole class	258	Pre	76(1)				
			Post	88(1)	0.48			
Edinburgh	Male	116	Pre	67(2)				
	Female	45	Pre	57(3)		10.6	0.005	
	Male	116	Post	86(1)	0.57			
	Female	45	Post	78(3)	0.50	7.7	0.013	-2.8
Hull	Male	40	Pre	62(3)				
	Female	6	Pre	43(6)		18.7	<0.001	
	Male	40	Post	77(3)	0.41			
	Female	6	Post	67(5)	0.43	10.1	<0.001	-8.6
Manchester	Male	198	Pre	79(1)				
	Female	60	Pre	66(2)		13.1	0.015	
	Male	198	Post	89(1)	0.49			
	Female	60	Post	82(2)	0.46	7.5	0.050	-5.6

Despite these reassuring normalised gains in improved conceptual understanding measured by the instrument, all three of the universities found a positive performance gender gap (positive G) in pre-instruction test results, as shown in Figure 3.14. This gender gap ranged from +10.6% to +18.7% prior to instruction. Results collected after a semester of teaching are encouraging, with a narrowing of the gap seen across all three institutions. However, the gap did persist and was statistically significant ($p<0.05$). The change in the gender gap between post-test and pre-test assessments ranged from -2.8%, at Edinburgh, to -8.6%, at Hull.

Despite each institution having different whole class mean scores on the pre-instruction assessment, it is clearly seen in Table 3.12 that female students entered the courses with lower FCI attainment. This was a result consistent across all

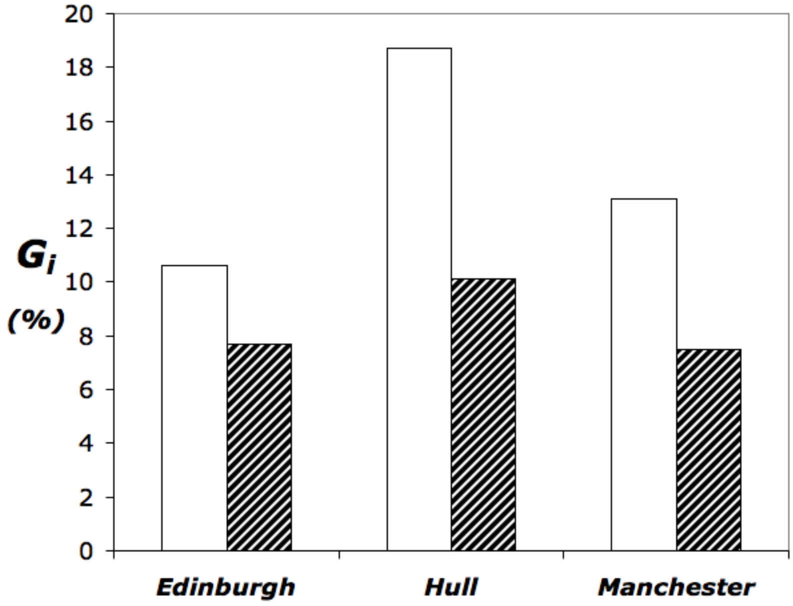


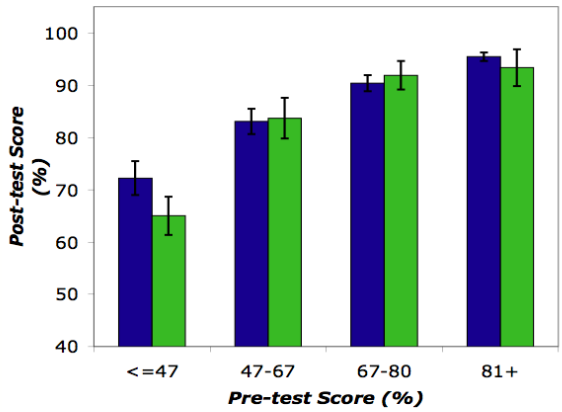
Figure 3.14: Percentage gender gap for University of Edinburgh, University of Hull and University of Manchester for pre-test (white) and post-test (hatched) in 2011-12.

three universities as well as with previous results found at University of Edinburgh which have been discussed earlier in this chapter. The difference between male and female scores was statistically significant at the 5% level both pre- and post-instruction in each of the three institutions.

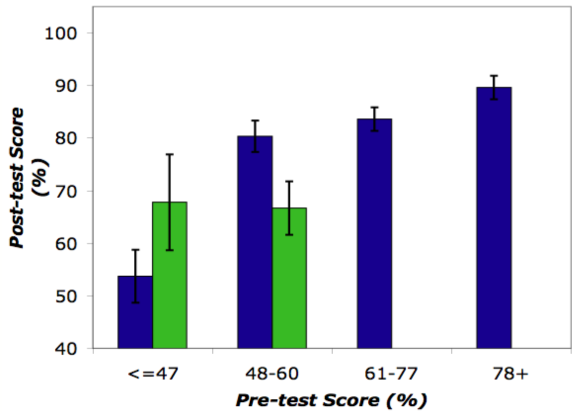
The distribution of students across pre-test scores and their outcome on the post-test was investigated. To do this the same methodology as in section 3.1.9 was completed. Each institutions' population was split into quartiles of approximately equal size on the basis of pre-instruction test performance. Each quartile was then separated into male and female subgroups and post-test scores calculated. Figure 3.15 illustrates the results for each university's gender-split quartile performance on the post-test.

Absolute number of students varied considerably between universities due to the overall cohort size. The total female population completing both rounds of the FCI assessment at University of Hull was particularly small (N_f=6). For the data from Hull, there were no female students in the top two quartiles and the small sample (particularly of female students) means it is unreasonable to try and draw conclusions from the distribution of the two gender populations.

(a)



(b)



(c)

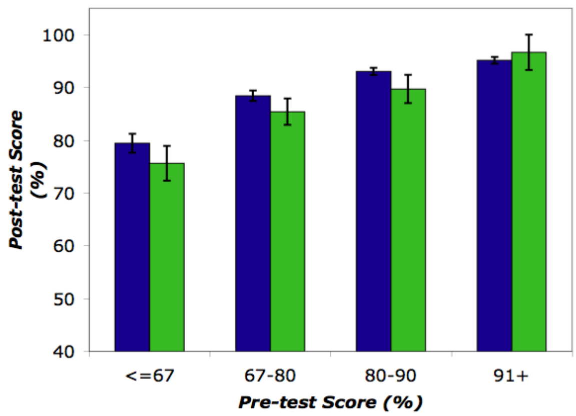


Figure 3.15: FCI post-instruction mean score for gender-split pre-instruction quartile groups. Male students are represented by blue bars, female by green. Data refer to students from (a) Edinburgh, (b) Hull and (c) Manchester. Error bars represent the standard error on the mean.

Both Edinburgh and Manchester had larger sample sizes, but results showed no statistically significant difference between post-test scores of the male and female cohorts within each quartile group. Although females performed slightly lower than males in the bottom quartile at both universities, this difference was not significant. Despite this, we do note that mean post-instruction test scores for the lowest scoring pre-instruction quartile barely exceeded the pre-test average for the whole cohort at the corresponding institution. On average students in the lowest quartile pre-instruction showed the lowest normalised gain post instruction.

By considering the proportions of male and female students in each of the pre-instruction bins (Table 3.13), a greater understanding of the gender gap can be gained. The fraction of male students in each of the four quartiles was approximately equal, and furthermore this was consistent for the male student cohorts from all three institutions. Males were distributed relatively evenly across the ability range upon entry to first year, prior to any instruction taking place. This was in direct contrast to the distribution of female students. The results from 2011-12, shown in Table 3.13, present a worrying picture about the distribution of females across the whole class performance. Across all three institutions, approximately half the female students in each of the institutions was in the lowest quartile prior to instruction. These results reflect those seen at Edinburgh over the five years prior to this data being collected and therefore show that this is not a feature unique to the University of Edinburgh.

Table 3.13: Fraction of male and female students in each quartile group of pre-instruction FCI scores for three UK universities. $N_{\mathrm{tot/M/F}}$ represents the number of students in total, those who are male and those who are female, respectively. $f_{\mathrm{M/F}}$ gives the proportion of male or female students, respectively, who are in each of the four quartile groups Q1 (highest) to Q4 (lowest) expressed as a fraction of the total number of male or female students in the cohort.

				Q1		Q2		Q3		Q4	
Institution	N_{tot}	N_{M}	N_{F}	f_{M}	f_{F}	f_{M}	f_{F}	f_{M}	f_{F}	f_{M}	f_{F}
Edinburgh	161	116	45	0.30	0.13	0.24	0.20	0.22	0.22	0.23	0.44
Hull	46	40	6	0.20	0.00	0.33	0.00	0.25	0.50	0.23	0.50
Manchester	258	198	60	0.21	0.03	0.30	0.20	0.29	0.27	0.21	0.50

An alternative way to analyse the data is to consider the transitions of students between quartiles from the start to the end of the semester. Of particular interest is the behaviour of those students who begin the course in the lowest ability quartile. Considering the Edinburgh data, we find that approximately 70% of students initially in the lowest quartile were also found in the lowest quartile on the post-instruction tests, with all of the remainder elevated to just the third quartile. For Manchester almost 60% of those students initially in the lowest quartile on the basis of pre-instruction FCI scores, remained there. This is of particular concern when we consider female students in this quartile. Approximately half of females start in the lowest quartile, the majority remain there, and for these students their post-instruction test performance remains, on average, the lowest of all eight sub-cohorts for the larger data sets from Edinburgh and Manchester.

3.3 Conceptual understanding in second year undergraduate physics

Having established a consistent gender difference in students' understanding of Newtonian force concepts, it was investigated whether such differences pertain only to first year courses or if gender discrepancies in understanding or performance persist once students have progressed into the second year of their degree program. It is important to note that almost all students taking the second year courses are doing so as a requirement for their final degree course (majors) and are therefore not choosing to take it as an outside subject. In this section the performance of male and female students in two second year courses will be discussed, with reference to data collected in the 2011-12 and 2012-13 academic years. Two diagnostic tests were used to probe student understanding as well as form a baseline from which any effects that may have resulted from the use of a values affirmation intervention could be measured.

3.3.1 Force and Motion Conceptual Evaluation (FMCE)

As described in Chapter 2, the Force and Motion Conceptual Evaluation (FMCE) assessment is a research-based assessment developed by Thornton and Sokoloff to measure understanding of Newtonian mechanics [36]. The diagnostic instrument contains 47 multiple-choice items. As with the FCI, the FMCE was presented to the whole class online using pre- and post-testing methodology. The choice of the FMCE for use with the second year cohort resulted from a decision not to further test first year students who were already participating in the FCI. Administering both the FCI and the FMCE to the same cohort would conflict with any pre- and post-testing carried out. The decision was taken not to use the FCI with second year students because they had already been exposed to the test twice (pre- and post-test) in their previous year of study and the high average post-test results suggested that there would be little room for improvement in student performance if used for a second year.

The FMCE was administered to the 2011-12 second year physics cohort at the University of Edinburgh in week one of the academic year as part of the 'Physics 2A' course. This course covers material from three areas of physics: Dynamics and Relativity, Waves and Geometric Optics. Students attend four one hour lectures per week in addition to a three hour computational and data analysis session and a two hour workshop in which they work on physics problems. The degree examination is worth 70% of their final course mark with data analysis contributing 20% and weekly assignments a further 10%. In 2011-12 this course had a cohort size of 117 students; 80 males and 37 females.

3.3.2 FMCE results from whole class analysis

In total 94 students participated in the pre-test for a response rate of 80%. This fell to a 52% response rate (61 students) for students completing the post-test. Students were informed that their best score from their two attempts of the FMCE diagnostic test would contribute to their written assessment mark, equivalent to one weekly assignment. The average pre-test score for the whole class was 81(2)% with a modal score of 44 out of 47, and average post-test score was 90(2)%, again with a modal score of 44.

Only 48 students completed both the pre- and post-tests to form the matched data set. Of these 29 were male and 19 were female. A Mann-Whitney U test concluded that there were no statistically significant differences between the male and female performance on the FMCE prior to instruction (p=0.228), although male students performed slightly higher than female students; 84(3)% and 77(4)% respectively. After 5 weeks of instruction both cohorts showed improvement in their understanding. Post-instruction male students scored 91(2)% and female students scored 88(3)%. Once again this difference was not significant (p=0.463). The average normalised gain for students was 0.33. Overall, males had a higher normalised gain (0.38) than females (0.25).

It was concluded that the difficulty level of the FMCE was too low for the students being tested and was more suitable for a first year cohort. Of the 94 students who completed the pre-instruction test, 73% achieved a score greater than 80%. These high pre-test scores undoubtedly had an influence on the low uptake by students on the second administration of the test. Figure 3.16 shows the distribution of pre-test and post-test scores as a function of gender. It is evident from the histograms that the majority of both male and female students performed highly in the pre-test leading to a ceiling effect in the post-instruction results.

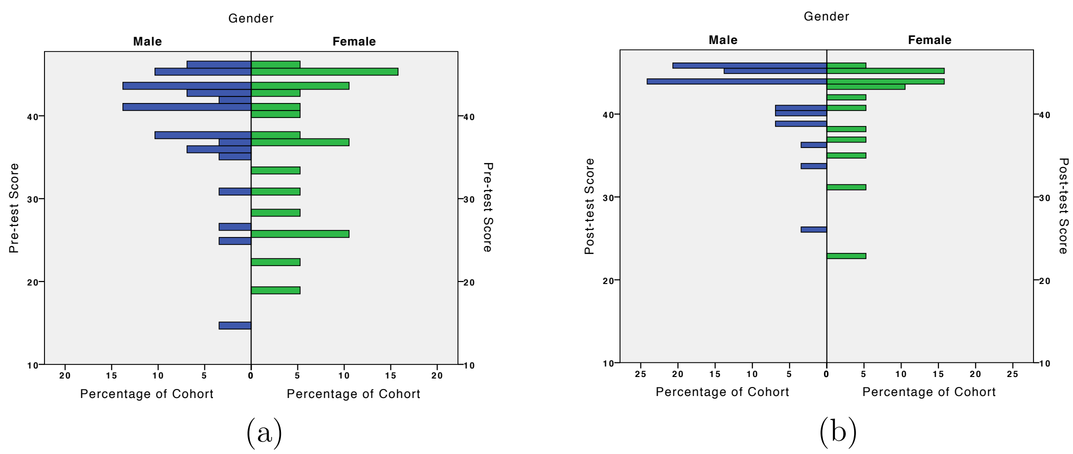


Figure 3.16: Histogram of gender distribution of FMCE (a) pre-test scores and (b) post-test scores for the 2011-12 cohort. N(males)=29 and N(females)=19.

Reported studies by Kost *et al* and Lauer *et al* both showed male students outperforming females on FMCE pre-test scores [3, 38]. Looking at post-test scores, Lauer *et al* reported an increase of 5.8% in the gender gap between the two administrations of the test. Unlike the results from Edinburgh, Kost *et al* found no change in the overall gender gap between the pre- and post-test as averaged over several semesters of the course. However, they did record variations in the gender gap for individual semesters.

3.3.3 Second year understanding of electricity and magnetism

The Brief Electricity and Magnetism Assessment (BEMA) multiple-choice test was developed in 1997 by Chabay and Sherwood [49] as an assessment tool to measure students' understanding and retention of concepts of electricity and magnetism. BEMA is a multiple-choice test consisting of 31 qualitative and quantitative questions, with up to ten possible answer choices, covering a broad range of concepts traditionally covered in an introductory electromagnetism course. A full copy of the BEMA assessment can be seen in Appendix D.

BEMA was administered to second year physics students at the University of Edinburgh in 2011-12 and 2012-13. In spring 2012 the BEMA was distributed during the second semester 'Physics 2B' course for second year physics students. This course comprised three sections; Properties of Matter, Electricity and Magnetism and Quantum Physics. The BEMA pre-test was administered to students in week 3 of semester, prior to teaching of the electricity and magnetism section of the course. The post-test BEMA was taken by students five weeks later (semester week 8). Changes in the curriculum in 2012-13 academic year resulted in the creation of a new 10-week course, 'Physics of Fields and Matter'. Students were again presented with the BEMA assessment prior to instruction (semester week 1) and at the very end of the course (semester week 10). In both years the test was administered online with a maximum time of 2 hours given to complete the test. The best of the students' two attempts (pre-test and post-test) contributed to their course grade, equivalent to one weekly hand-in assignment (approximately 2% of their final course mark).

3.3.4 Whole class BEMA results

Only matched student responses were included in the analysis so that changes in individual students' performance between pre- and post-tests could be recorded. This led to 54% of students being included in 2011-12 and 67% in 2012-13 analysis respectively. Table 3.14 shows the percentage of each cohort included in the matched data as a function of gender. The percentage of matched responses was 14% to 18% higher for female students than for male students.

Despite the structure of the second year electromagnetism course undergoing

Table 3.14: Number and proportion of second year students completing both pre-test and post-test administrations of BEMA.

			Male Students		Female Students	
Year	N cohort	% cohort	N cohort	% cohort	N cohort	% cohort
2011-12	47	54%	27	49%	20	63%
2012-13	79	67%	57	63%	22	81%

a significant change between the 2011-12 and 2012-13 academic years, the prior level of knowledge, as well as the material covered over the course of the electromagnetism course, remained fairly consistent between the two years. This was further reinforced by the consistency of the pre-test scores over both years. There existed no statistically significant difference in pre-test scores between the 2011-12 and 2012-13 cohorts. On average students scored 46(3)% and 45(2)% on the pre-test in 2011-12 and 2012-13 respectively. This is a higher pre-test result than that seen in other published studies which find the average BEMA pre-test score to be approximately 25% [101, 122, 123, 124]. In the case of the data presented in this section, results from both years in which the BEMA was administered have been combined in order to create one data set.

As can be seen in Figure 3.17, there existed a large range of scores pre-test, with a standard deviation of 17.7%. This wide spread of results persisted post-instruction, although there was evidence indicating that students in both academic years improved greatly after the teaching period. Once again there did not exist any significant difference between results from the two academic years, with average post-test scores increasing to 68(3)% in 2011-12 and 67(2)% in 2012-13. These results are slightly higher than those presented in previous studies [101, 124].

3.3.5 Gender analysis of BEMA results

In order to determine whether there were any differences between male and female students' understanding of electricity and magnetism, the results from the combined 2011-13 data set were analysed to find the average score for each gender, as shown in Figure 3.18. The results prior to any instruction showed no significant differences between genders, with male students scoring on average

(a)

(b)

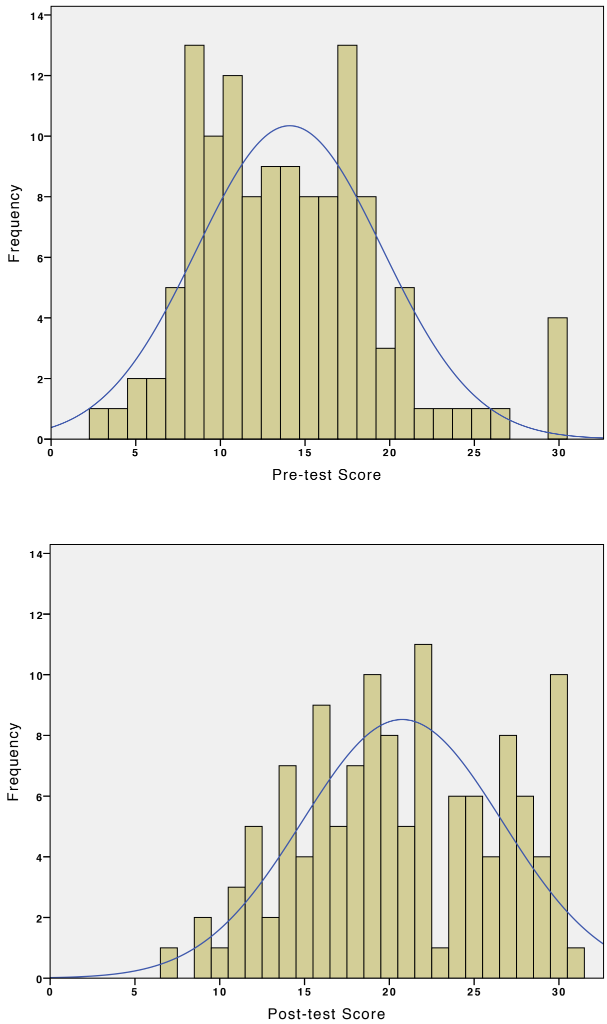


Figure 3.17: Histogram of (a) pre-test and (b) post-test BEMA scores from the combined data set of 2011-12 and 2012-13. N(2011-13)=126. Curve indicates a normal distribution.

45(2)% compared to an average of 46(3)% for female students. Both males and females on the course demonstrated an improved understanding of electricity and magnetism after one semester. Male students reached an average score of 69(2)% and female students a score of 63(3)% as measured by the post-test.

Other studies have observed the existence of a gender gap of the BEMA assessment, with the gender gap increasing post-test [57]. The lack of a statistical difference between males and females in our population is in contrast to results found during the 2007 Fall semester at the University of Colorado [124]. In their study a statistical difference was found in the pre-test scores, with males

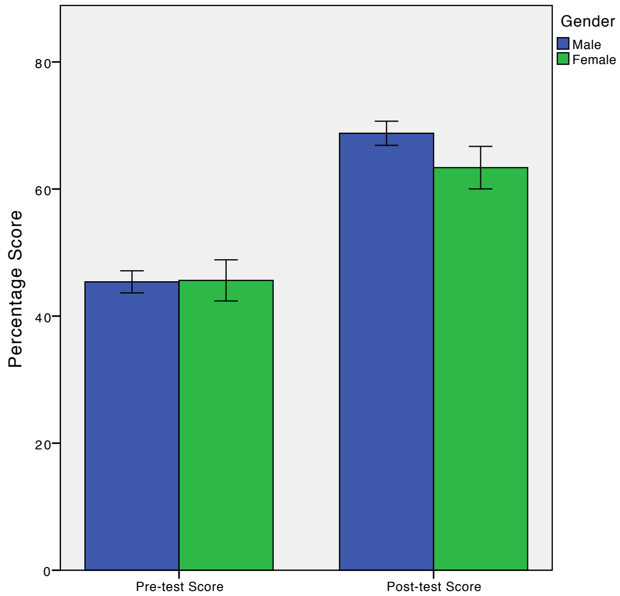


Figure 3.18: Graph showing mean (a) pre-test and (b) post-test BEMA scores for male and female students between 2011-13. N=126. Error bars represent the standard error on the mean.

scoring more highly than females. In fact, this gender gap increased in the post-instruction results. Previously reported results from a multi-semester study at the University of Colorado indicated that in five of the eight semesters tested there existed no significant gender differences at the 95% confidence level between male and female pre-test BEMA scores [125]. The statistically significant gender gaps that existed in the remaining semesters were much smaller than those typically seen on the FCI or FMCE (males scoring between 2.6-3.6% higher than females). Conversely, the post-test results showed a significant gender gap in all semesters.

These results differ from those seen in the Force Concept Inventory which, as discussed earlier in this chapter, showed a significant and consistent gender difference between male and female students both prior to and after instruction. The absence of a measurable gender difference in conceptual understanding in the second year course is perhaps suggestive of students' exposure to certain physics concepts prior to university. Students have minimal exposure to concepts of electricity or magnetism prior to the instruction they receive during this second year course, either in their first year as undergraduates or indeed at secondary school. In contrast, Newtonian concepts of force and motion are a primary focus of physics at secondary school level. These results may imply that a lack of prior teaching of electromagnetism has meant that they are less likely to have formed preconceived ideas about this topic prior to this course. Both genders are receiving their initial teaching of the subject at the same time, through the same teaching methods.

3.3.6 BEMA question by question analysis

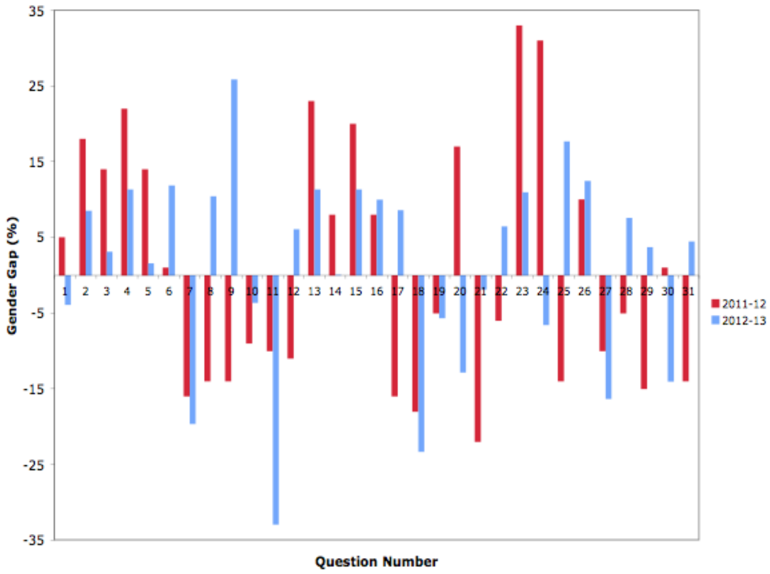
Despite no gender difference in the overall test scores of students, analysis of individual items on the BEMA test showed large variations in the percentages of males and females answering each question correctly. Figure 3.19 shows the percentage gender gap[2] for each question on the diagnostic test for both 2011-12 and 2012-13 academic years in the pre- and post-tests.

Although the second year course focuses on electricity and magnetism, students are not given any formal teaching regarding electric circuits which are tested in seven of the test questions. Despite this, it is expected that students would be familiar with the content of these questions from previous study at school level. When interpreting the statistical significance between gender performance, one must be aware of the relatively small sample sizes, particularly for females, in each of the year groups discussed.

Comparing results from the 2011-12 academic year with those from 2012-13 we see that, prior to instruction, males performed better on 15 of the questions in 2011-12 and 19 of the questions in 2012-13. Depending on the test item, there was a large range in the measured percentage gender difference of the number of students answering correctly; between -33% and +33%. The sign of the gender gap is, however, not consistent for each question between the two years; in 2011-12 males sometimes performed better on a specific question and in 2012-13 females sometimes performed better on the same question and vice versa. The sign of the gender gap was the same for both years in 17 of the 31 test items. The large variation in the direction of the gender gap between year groups makes it unclear whether these results are highlighting a true gender difference or are the result of fluctuations or 'guessing' due to a lack of knowledge amongst students. In this section the difference in male and female performance on a selection of

[2]The percentage gender gap has been defined as the percentage of male students answering the question correctly minus the percentage of female students answering the question correctly. A positive percentage gender difference indicates that a higher proportion of male students identified the correct answer than female students.

(a)



(b)

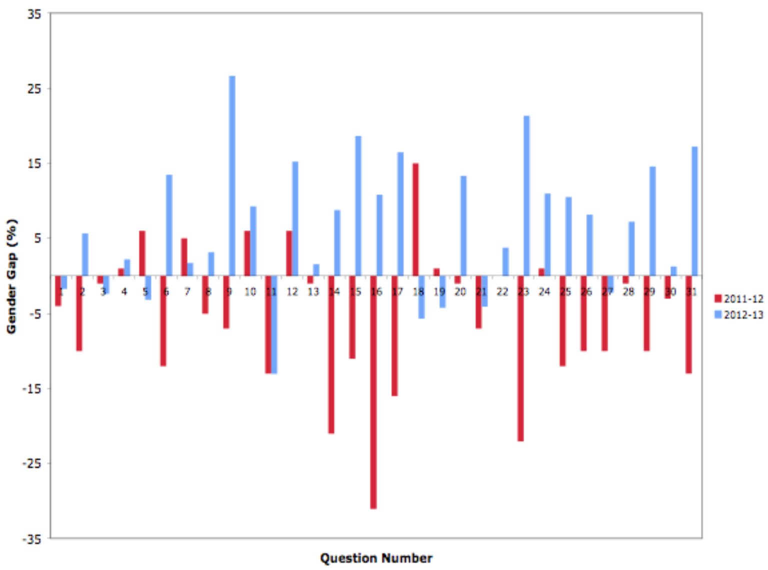
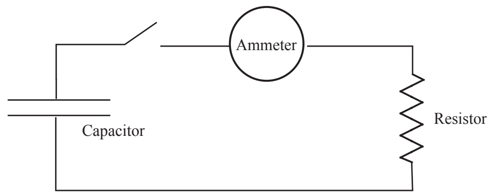


Figure 3.19: Percentage gender gap of students answering each question of BEMA correctly (a) pre-test and (b) post-test. The percentage gender gap has been taken as the percentage of male students answering the question correctly minus the percentage of female students answering the question correctly.

items from the BEMA assessment will be presented in more detail, with reference to questions showing contrasting changes in gender performance gap after one semester of teaching.

Question 13

Figure 3.20 shows Question 13 of the BEMA test. Students are shown a diagram of a circuit containing an ammeter, a capacitor and a resistor in series with a switch and are asked how the current in the ammeter will behave if the switch is closed. From 2011-13 a total of 74% of students answered this question correctly in the pre-test. In both years the gender gap was positive, indicating that a higher proportion of male students answered this question correctly. The gender gaps was 23% in 2011-12 and 11% in 2012-13. After a few weeks of teaching the difference between male and female performance was almost entirely eliminated on this item, with measured differences of only -1% and 2% in 2011-12 and 2012-13 respectively.



▶ Q13: The capacitor is originally charged. How does the current I in the ammeter behave as a function of time after the switch is closed?

(a) $I = 0$ always

(b) $I = \text{constant} \neq 0$

(c) I increases, then is constant.

(d) I instantly jumps up, then slowly decreases.

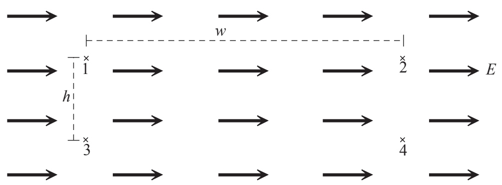
(e) None of the above.

Figure 3.20: Question 13 from BEMA. See Appendix D for complete version of test [49].

Question 14

Question 14, depicted in Figure 3.21, was the first of three questions pertaining to the same diagram showing a uniform electric field. Students were asked to identify the potential difference between two points in this field. Approximately 47% of students answered this question correctly at the beginning of the semester, increasing to 67% after teaching. This question demonstrated almost no pre-test gender difference between students in 2012-13 and only 8% more males than females answered it correctly in 2011-12. In 2011-12 the gender gap increased greatly to -21%. The negative percentage gender gap indicated that females outperformed males. The 2012-13 cohort also showed a growth in the gender gap to 9%.

In a certain region of space there is a uniform electric field of magnitude E:



Choose from the following possible values to answer the three questions below:

(a) $+Ew$

(b) $-Ew$

(c) $+Eh$

(d) $-Eh$

(e) $+E\sqrt{(h^2 + w^2)}$

(f) $-E\sqrt{(h^2 + w^2)}$

(g) zero

▶ Q14: The potential difference $V_2 - V_1 = ?$

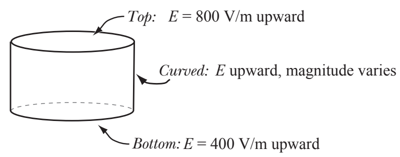
Figure 3.21: Question 14 from BEMA. See Appendix D for complete version of test [49].

Question 18

Figure 3.22 shows Question 18 of the BEMA test which asks students to identify the charge enclosed in a cylinder. Students choose one of five multiple choice statements to answer the question. This question was answered correctly by only 44% of students in the pre-test. In both 2011-12 and 2012-13 females performed more highly than males (a negative percentage gender difference). In 2011-12 18% more females answered this question correctly than males and in 2012-13 23% more females answered this correctly. Interestingly, the gender gap narrowed after teaching to -6% in 2012-13, but in the previous year, males outperformed females by 15% post-instruction.

Analysis of student performance on individual items of the BEMA test suggests a lack of a clear pattern in the change in the gender gap over time. A large proportion of questions showed a different sign of gender gap (positive or negative) depending on the academic year studied. As a result, it is difficult to draw definitive conclusions about any potential gender preferences for individual test items from these two years of data collection. Irrespective of this, it can be seen that an almost identical overall performance on the diagnostic test by males and females is not reflected by identical performance on individual items on the test. Male students are outperforming females on some questions and females are outperforming males on other questions. Collection of data from future year

Here is a cylinder on whose surfaces there is an electric field whose direction is vertically upward, but whose magnitude varies as shown.



▶ Q18: The cylinder encloses

(a) no net charge.

(b) net positive charge.

(c) net negative charge.

(d) There is not enough information available to determine whether or not there is net charge inside the cylinder.

Figure 3.22: Question 18 from BEMA. See Appendix D for complete version of test [49].

groups may help to determine if the lack of gender gap on the test overall is a continuing trend.

3.3.7 Values Affirmation study

A Values Affirmation exercise was administered in second year physics courses in 2011-12 to investigate its effect on whole class performance as well as acting as an intervention to potentially improve female performance and reduce the overall gender performance gap. As discussed in Section 1.1.1, Miyake *et al* investigated the use of a short writing exercise, in which students wrote about their personal values, on improving the performance of female students on the Force and Motion Conceptual Evaluation (FMCE) [36, 37]. It was hypothesised that many females experience additional pressure not to affirm stereotypes which suggests that females are more likely to perform negatively compared to males in sciences [37]. By engaging students and providing them with the opportunity to affirm their own personal values in a non-threatening environment, it is suggested that they build up personal coping mechanisms.

Previous studies have suggested that students with a high level of self efficacy or belief in their own learning capabilities are more likely to show higher gains in inquiry-based assessments. A study by Cavallo *et al* found that self-efficacy was a positive predictor for both male and female students' conceptual understanding and overall course achievement [126]. Their study of an introductory physics course for biology students also indicated that females had a significantly lower level of physics self efficacy throughout the course compared to males. Similarly, a study by Shaw found significant differences between the self-efficacy of males and females in non science major classes [127]. These gender differences may contribute to observed differences in academic performance both at secondary school and university.

The affirmation exercise was first used at Edinburgh in conjunction with the FMCE. As mentioned previously, it was decided to administer this test to second year students so as not to influence results collected from first year students completing the FCI. These results proved inconclusive due to the difficulty level of the FMCE proving too low, with almost 75% of students achieving scores of at least 80% pre-instruction. The study was subsequently conducted during the second semester 'Physics 2B' course. The BEMA test was used to establish a baseline for students' prior performance and administered again after completion of this section of the course to measure student improvement.

3.3.8 Methodology

The Values Affirmation was disseminated during the tutorial session directly following the closing of the BEMA pre-test. Students were randomly assigned to either the 'control' group or the 'values affirmation group'. All students sitting at the same tutorial table were assigned to the same group to avoid students recognising that there were two different forms of exercise.

At the start of the tutorial students were given a brief introduction informing them that the purpose of the exercise was to encourage effective writing and communication skills which become increasingly important as they progress through their degree. The head of the physics tutorial workshop introduced the exercise and told the students that they would not be writing about physics, but instead would be discussing what values they feel are important to others and to themselves. All teaching assistants aimed to minimise the possibility of students discovering the difference between the control and affirmation exercises. Students were asked to complete the exercise individually without conferring with their neighbours. Participants were also told that they would receive short feedback on their individual writing exercises in a future tutorial workshop. They were asked to write their matriculation numbers on their work so that BEMA results could be correlated with students in either the 'control' or 'values affirmation' groups, although it was made clear that no one outside this study would see their assignment and no course credit would be given for participation.

The first page of the assignment listed the twelve personal values used by the original authors [37]. Those students assigned to the 'value affirmation' group were instructed to circle two or three of the values they felt were most important to them. Similarly, students in the 'control' group were instructed to circle two or three values least important to them. The options were as follows:

- Relationships with family and friends
- Government or politics
- Independence
- Learning and gaining knowledge
- Athletic ability
- Belonging to a social group (such as your community, racial group or school club)
- Music
- Career
- Spiritual or religious values
- Sense of humour
- Art
- Creativity

The second part of the exercise instructed students to describe in a few sentences either why the values they had previously selected were important to them (affirmation group) or why they might be important to someone else (control group). Finally, students were again asked to look at the values they selected. They were instructed to list the top reasons why these were important to them or others.

3.3.9 Values Affirmation results and correlation with BEMA

Only 23 students from the 2011-12 matched BEMA data set completed the values affirmation exercise. Of these students, 11 were male and 12 were female. As a consequence of the small number of participants, it is almost impossible to derive a clear conclusion about the effect of the affirmation intervention on student performance. These 23 students were representatives of both the 'control' and 'values affirmation' groups; 14 control students and 9 values affirmation students. Both these subgroups performed equally in the pre-test, as shown in Table 3.15. Although the control group achieved a higher post-test score, the large standard errors on the means resulted in no statistical difference between cohorts. The very small numbers of students in each gender subgroup, in conjunction with the lack of gender difference in the electricity and magnetism assessment, either prior to or post instruction, makes it difficult to conclude if this had a preferential effect on a specific gender of student. Further data collection is required in order to establish if this intervention could have a desirable effect on closing observed gender gaps.

Table 3.15: Results from BEMA diagnostic test for students who completed the Values Affirmation exercise. Numbers in brackets represent the standard error on the mean. N(male)=11 and N(female)=12.

Test Group		N	Pre-test (%)	Post-test (%)
Control	Total	14	43(4)	71(6)
	Males	6	47(5)	74(9)
	Females	8	40(5)	69(8)
Values Affirmation	Total	9	43(8)	60(9)
	Males	5	43(6)	68(12)
	Females	4	44(19)	51(16)

Miyake *et al* found that applying the values affirmation exercise with students from an introductory physics course resulted in a decreased gender gap, as measured by the FMCE, for those students who completed the writing exercise [37]. Students who were in the control group had a larger gender gap. A later replication of this study with the same instructor and course did not result in an improvement in the gender gap [39]. Results did suggest that females who completed the self-affirmation exercise outperformed females from the control group on course examinations. This may suggest that similar interventions are an area ripe for further investigation and may provide an opportunity for improving the gender disparity in science courses. It is also important to be aware of potential consequences of students' attitudes and self-efficacy. Further discussion of students attitudes to studying and learning physics will be presented in Chapter 7.

3.4 Chapter discussion and summary

In this chapter diagnostic tests have been extensively employed to investigate the performance of different first and second year undergraduate populations on conceptual understanding assessments. As well as results from whole class cohorts, analysis has been undertaken to compare physics majors and non-majors as well as male and female students. The differences between these populations have been measured to gain a clearer understanding of the existence of such performance gaps as well as how the difference between male and female performance changes over the course of their first two years of study.

Results from seven consecutive years of implementing the FCI in first year physics classes have shown evidence for a clear and consistent gender gap at the point of entry to university. In each academic year males entered the course with a greater understanding of Newtonian concepts of force and motion than females, as measured by the concept inventory. This difference cannot be explained by a disparity in the school leaving qualifications of incoming students, who have all achieved the necessary school results. In fact there is evidence to show that female students outperform males in school-leaving exams, both in physics and other subjects [116]. Hazari *et al* considered the impact of US high school education on student performance and found that female students entering university physics courses had a statistically stronger background in most subjects, although this difference was not significant for physics, but nevertheless performed at a lower level in the introductory courses than male students with the same background [128]. Following a semester of teaching at Edinburgh, in which the level of interactive engagement methods implemented has increased from year to year, this gender difference narrowed. Although the existence of a gender gap has been widely reported in previous studies, these results suggest a worrying pattern [55, 57, 62]. The statistically significant gender gap in our first year course persisted post-instruction and the difference between male and female scores was not entirely eliminated, with females still under-performing compared to males. Results from a calculation of normalised gains for the whole cohort did suggest that students showed an encouragingly high level of improvement over the course. In five of the seven years females had a higher normalised gain than males.

The relationship between pre- and post-instruction scores showed that, on average, students with similar FCI scores at the start of the academic year tended to have comparable post-test scores. Although there was an overall gender difference in conceptual understanding, no indication of a difference between males and females scores in any of the four performance quartiles between 2006-10 was found. Data from 2011-13 showed that female students in the lowest performing pre-test quartile significantly underperformed in the post-test compared to male students. Interestingly, there was an observed discrepancy in the distribution of students across the quartiles. Females were disproportionately represented in the lowest performance quartile. What was particularly concerning was that approximately 50% of the female cohort who completed both the pre- and post-tests lay in the lowest quartile bin. Conversely, the male population was distributed more evenly across the quartiles. Results suggested that a large proportion of those who started the semester in the lowest quartile remained in this quartile after a semester of teaching.

A comparison study showed that these results are reproduced at other UK institutions and are not unique to the University of Edinburgh. Despite differences in the educational contexts at the University of Hull and the University of Manchester, as well as variations in method of delivery of the FCI to students, all three universities showed high learning gains and a positive gender gap between males and females before instruction. Similar to the results seen at the University of Edinburgh, the gender gap narrowed after one semester, but remained statistically significant. Taking into account the small number of female participants at the University of Hull, again it was seen that at all three institutions around half of females were in the lowest ability quartile prior to teaching.

Looking at physics majors and non-majors, large differences in students' ability in conceptual understanding tests of Newtonian concepts of force were seen. Physics majors entered with a significantly higher level of conceptual understanding than physics non-majors, despite both cohorts holding the same entry qualifications. One explanation for these observed differences may be that students' perceptions of their studies and course choices may influence their motivation and thus their overall performance [129]. Non-majors may believe that they should focus their studies on their compulsory degree courses and may therefore view their 'outside' subjects as less important. The non-majors considered in this thesis are dissimilar to those in North American studies. Non-major students at the University of Edinburgh are required to have the same entry qualifications as physics majors, the only distinguishing feature being that they have not chosen to enrol on the physics degree programme. Both groups improved over the semester but the performance gap remained significant.

Notable differences were also seen when comparing male and female physics majors and non-majors. Male majors consistently outperformed other subgroups in the pre-test FCI in each of the seven years analysed. In the 2006-10 data presented in this chapter, a narrowing of the gap between male majors and non-majors was seen, although the gap remained statistically significant post-test. This is in contrast to results from the final two years of data where the difference between male majors and non-majors was no longer statistically significant after a semester of teaching. The gender gap between male and female major students was also evident both pre- and post-instruction, with males performing much higher than females, with the exception of 2009-10 when females and males performed equally in the FCI post-test. Interestingly, there did not exist a statistical difference between female majors and non-majors in the 2006-10 combined cohort. Male non-majors did significantly outperform female non-majors both at the start of the semester and after ten weeks of teaching. Large variations in performance were witnessed from year to year. This was particularly the case for female non-majors, who in 2010-11 outperformed both male non-majors and female majors. There is reason to believe that the inconsistencies between year groups may be an effect of the changing composition of the course cohort. Changes in the recruitment method by the university existed. In 2010-11 a selection process was employed which may have had an effect on the incoming cohort of students, with the introduction of higher entry qualifications for physics students. The gender profiles of each year group also showed variations. In particular, the number of female non-majors in the final three years of data was much larger than that of female majors (Table 3.4). These difference have been taken into account when drawing conclusions from this analysis. In addition to this we cannot directly compare FCI data between all year groups, since a different version of the test was used.

As well as establishing differences in conceptual understanding in first year physics cohorts, it was investigated whether evidence of similar performance profiles existed in second year courses. Through the administration of two diagnostic tests, one focusing on concepts of force and motion and one relating to electricity and magnetism, it was seen that the tests did not elicit any underlying gender differences between cohorts. Results collected from the distribution of the FMCE in semester one suggested that the level of difficulty of the diagnostic assessment was too low for the prior knowledge of the students and offered very limited scope for students to improve after teaching. It is therefore difficult to conclude whether the lack of gender difference observed is an underlying characteristic of the cohort or the result of prior first year learning. Male and females students both showed marked improvement on the BEMA assessment after one semester. The level of understanding measured by the BEMA pre-test scores was consistent, if not slightly higher, than that witnessed in previous studies [57, 124]. The lack of a significant gender difference either pre- or post-test suggests that both groups of students have similar levels of learning. The absence of a statistical difference between male and female performance on the BEMA test may suggest a further contextual dimension to the gender performance problem. Unlike Newtonian mechanics which forms an integral part of the secondary school syllabus and is the first physics topic to which most students are exposed, electricity and magnetism are topics to which students have been given little prior exposure, either during their secondary education or during the first year of undergraduate physics teaching. Both genders are therefore introduced to this topic at the same time.

One explanation for the difference between the observed gender results from the first year FCI and second year BEMA results may be that females are more likely to embrace a 'common sense' belief [103]. Research has suggested that females are more inclined to try to relate physics concepts to 'real world' situations, and in this process develop common misconceptions [52, 130, 131]. These misconceptions could then be manifested in results of the FCI. Secondary school courses often use everyday examples to explain physics concepts, something that is not relied upon when teaching electricity and magnetism. Although most published studies report the existence of a gender gap on the BEMA test, in favour of male students, the magnitude of this gap is, on average, much lower than that seen on tests of Newtonian mechanics such as the FCI or FMCE [57]. It it however unclear why female students might retain this world view to a greater extent than males. These results may suggest that the gender gap is specific to Newtonian mechanics, a topic more befitting to personal experiences.

The prevalence of common misconceptions is a concern. The work presented in this chapter raises the questions of what misconceptions are held by students and whether female or male students are more likely to embrace these. An analysis of these misconceptions will be discussed in more detail in Chapters 4 and 5.

Chapter 4

Question by Question Analysis of Student Misconceptions

Students often enter introductory physics courses with well established misconceptions and common-sense beliefs about certain physics concepts, particularly force and motion [52, 132]. Conceptual tests such as the Force Concept Inventory have been designed and employed to measure the prevalence of these misconceptions amongst student populations. In this chapter quantitative data collected for individual items on the FCI are reported, followed by a more in-depth study examining student response profiles.

When looking at the performance of first year undergraduate students at the University of Edinburgh in Chapter 3, it was seen that students had a statistically significant increase in their conceptual understanding of Newtonian mechanics following one semester of teaching and this was consistent with other studies. Despite this increase in whole class performance, the significant difference between male and female students which existed at the point of entry to university persisted at the end of the semester, with females still under-performing compared to males. This observed gender gap was reflected in results collected from two other UK universities [118]. The consistency of these gender differences, which appear to be independent of the delivery method of the assessment and teaching methods used in each of the universities, gives reason to question whether the discrepancies arise from males performing more highly on specific inventory questions or whether they originate from males outperforming across the whole test.

As discussed in section 1.2.4, it has been noted that student performance on assessment questions can vary widely depending on both the concept being tested and the context in which it is presented. In a study by Kohl *et al* students presented with isomorphic questions showed different levels of performance depending on the representation of the concept being tested [133]. Meltzer *et al* found that student performance on questions testing the same concept varied depending on the question representation, with some students' answers showing inconsistencies between different representations [134]. Noticeable differences were seen between female and male performance on graphical questions on a Coulomb quiz, with females performing more poorly than males [134].

With respect to gender differences, a study at the University of Wisconsin by McCullough investigated the effects of changing the contextual representations of FCI questions from what was considered stereotypically male contexts to overtly female contexts [75]. Although female students showed no significant changes in performance between the two tests, males taking the test based on 'female contexts' showed a drop in performance. Results of this study however were inconclusive due to very low overall FCI scores by the whole cohort. The existence of observed gender differences at Edinburgh was not wholly unexpected with previous studies showing changes in the performance gap between male and female students [55, 62]. There has however been less study into identifying specific differences in misconceptions between males and females.

Multiple choice tests force students to choose between the correct answer and distractors, created specifically to highlight common misconceptions surrounding physics concepts [135]. Presented in this chapter is a detailed examination of students' performance on individual items from the FCI. The following sections build on the results obtained from the analysis of overall FCI scores in Chapter 3, looking more specifically at the multiple choice response profiles for each test item. Proportions of male and female students choosing different distractors are explored in order to investigate gender differences in levels of conceptual understanding.

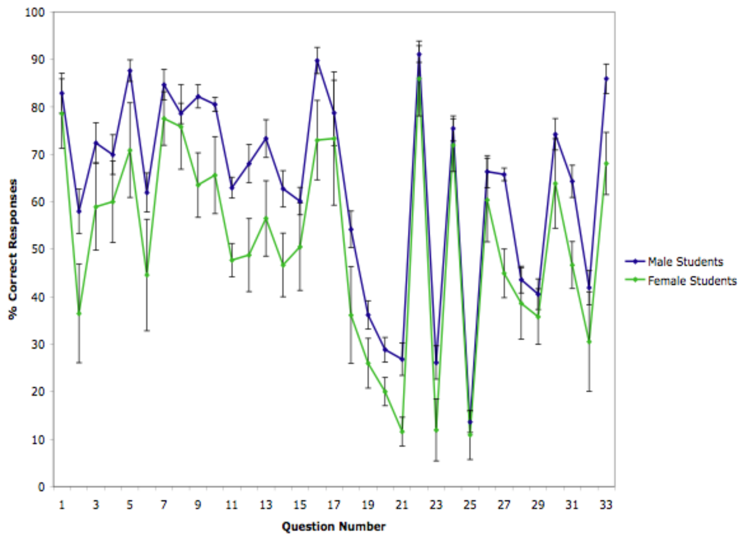
4.1 Question by question analysis of the FCI

In order to fully understand the origins of the observed gender gaps discussed in Chapter 3, further research is required to identify potential differences between male and female students' understanding of different physics concepts. Pre-test results from the implementation of the FCI in our first year physics course showed a wide range of student scores. This spread in performance was reflected in the variation in number of correct responses collected depending on the specific test item analysed. In this section an analysis of gender differences on individual items of the FCI will be discussed with respect to first year undergraduate students between 2006-13. For each of the questions on the FCI the number and percentage of students who answered each question correctly was calculated. The data was subsequently split by gender to look for any differences between males and females. Because we were interested in performance on individual test items, unmatched data was used in the analysis.

Looking first at data collected in the first year introductory physics course at the University of Edinburgh between 2006-10, combined results for the pre-test, shown in Figure 4.1 (a), illustrated that there was a lower percentage of correct responses for female students compared to male students for all 33 questions. This is in agreement with results from a calculus-based physics course at the University of Minnesota [62]. There was considerable variation in the percentage of correct responses depending on the question answered, indicating a greater or lesser understanding of certain concepts. In the 2006-10 data set only 8 of the 33 test questions did not show a statistical difference at the 95% confidence level between the percentage of male and female students answering correctly in the pre-test. This analysis was repeated for the post-test results, and, in addition to an overall improvement in student performance, a reduction in the gender gap compared to that in the pre-test was seen (Figure 4.1 (b)). Females still had a lower percentage of correct responses in all questions, but fewer questions showed a statistically significant difference between the populations. In the post-test results 19 questions showed no significant gender differences, again illustrating the narrowing of the overall gender achievement gap.

Figures 4.2 (a) and (b) show the results from the equivalent pre- and post-test analysis for the combined 2011-12 and 2012-13 year groups. Once again, males had a higher percentage answering each of the questions correctly compared to

(a)



(b)

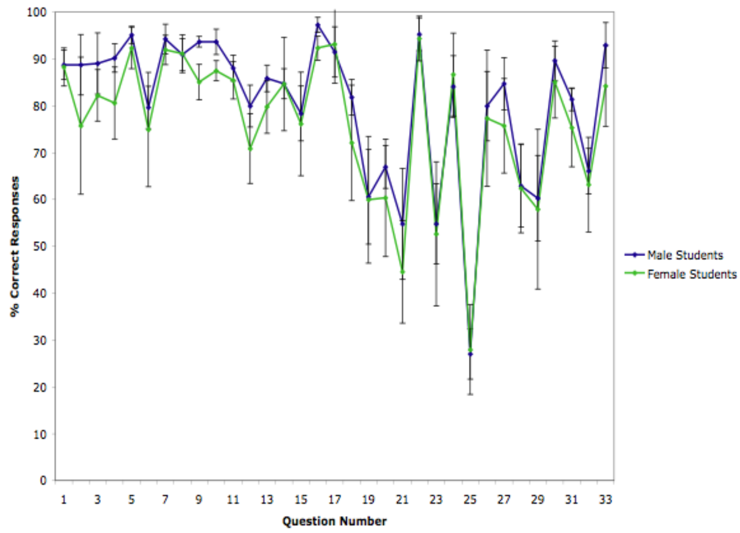
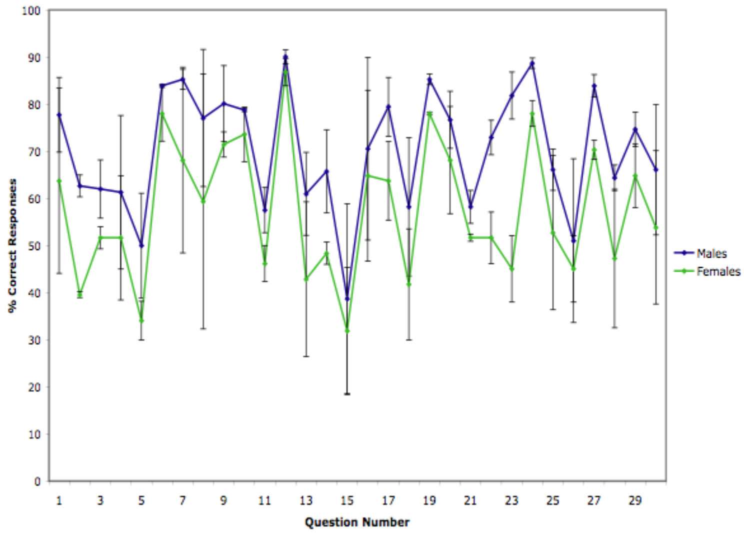


Figure 4.1: Percentage of correct responses for each question on the FCI_{ext} (a) pre-test and (b) post-test for 2006-10 combined data as a function of gender. Error bars represent the standard deviations on the means over five years. N(males)=692 and N(females)=247.

females. There was also a shift to higher percentages of correct responses in the post-test results, as reflected in the overall scores on the FCI at the end of semester one discussed in the previous chapter. Results from the combined pre-test scores indicated that 13 of the 30 questions in the FCI demonstrated no statistically significant difference between genders. This increased to 17 questions

(a)



(b)

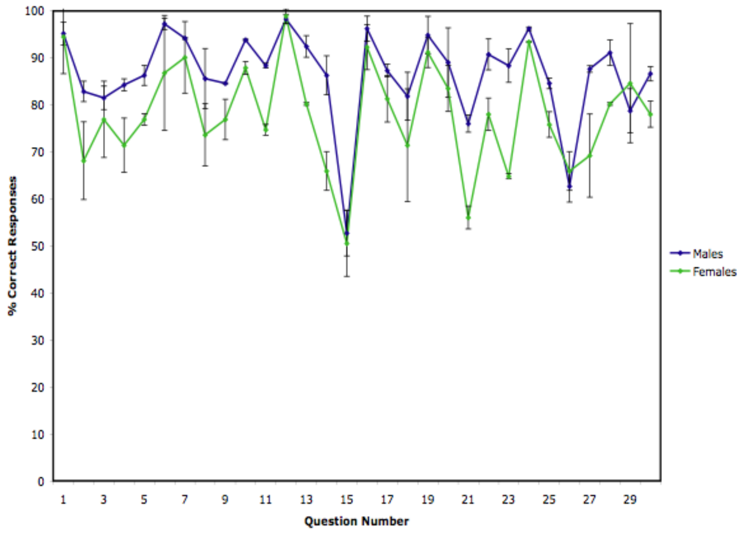


Figure 4.2: Percentage of correct responses for each question on the FCI (a) pre-test and (b) post-test for 2011-13 combined data as a function of gender. Error bars represent the standard deviations on the means over two years. N(males)=292 and N(females)=91.

in the post-test.

When comparing the 2006-10 and 2011-13 data sets, it is important to remember that the questions are not identical (because two different versions of the FCI were used) and, those that are approximately the same contain subtle differences in representation or context and differ in the order in which they are presented. We can however note that, on both tests, there exist some items that cause students more difficulty than others. Identifying such questions, and the physics concepts they test, can be beneficial in helping improve both the gender disparity and overall student conceptual understanding.

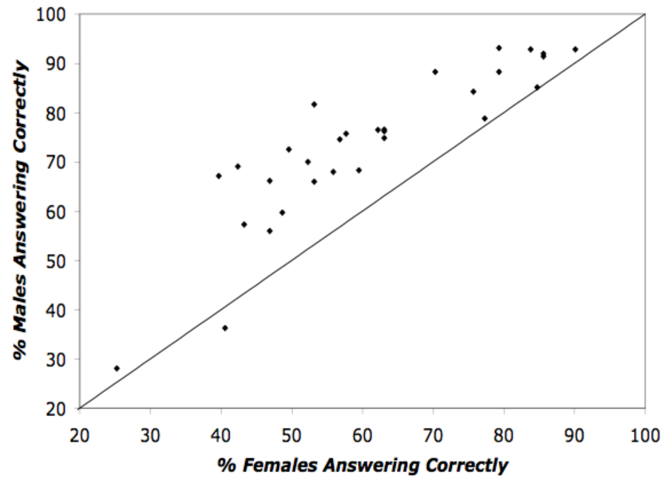
4.1.1 Item analysis from comparisons at three UK universities

Work carried out in conjunction with the Universities of Hull and Manchester in 2011-12 enabled the collection of FCI test data from three different first year physics cohorts. As discussed in Chapter 3, an overall gender difference was found on the FCI at each university, both prior to and after instruction. Pre- and post-test data from these three universities were examined to investigate the changes in the gender gap on individual items on the FCI over the course of one semester of teaching [118]. Findings from this project showed large variations in gender performance on individual test items.

Figures 4.3 (a) and (b) illustrate the gender gap in performance, for the combined Edinburgh, Hull and Manchester data, on each of the 30 FCI items for the pre- and post-test, presented as a plot of the proportion of male students getting an individual item correct against the corresponding proportion of female students who do likewise. Previous results, showing an overall gender disparity, in which males significantly outperformed females, lead us to believe that we should not expect the line of unit slope to represent a line of best fit to the data. For a data point to reside on this line there must be an equal proportion of male and female students answering that question correctly. It can be seen that the majority of the data points in fact lie above the line of unit slope, indicating that a higher proportion of males answered that question correctly than females. This was not the case for Question 26 where a higher proportion of females responded correctly than males in both the pre-test and post-test. In addition to this, results from Question 19 in the pre-test indicated that 85% of both the male and female populations identified the correct answer and therefore this question lay on the line of unit slope.

These findings support our results from the University of Edinburgh between 2006-10 and 2011-13, discussed in the previous section, which showed a larger fraction of male students getting a given item correct compared to females

(a)



(b)

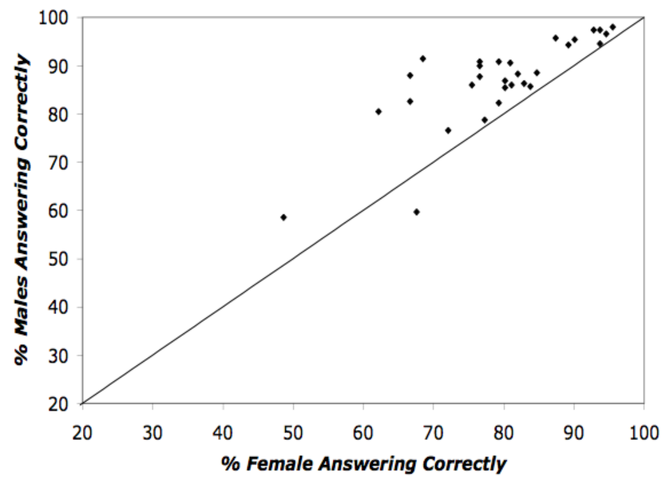


Figure 4.3: Percentage of male students versus the percentage of females students answering each item on the FCI correctly on the (a) pre- and (b) post-instruction tests. Data represented here is combined data from the University of Edinburgh, University of Hull and University of Manchester for the 2011-12 academic year. The line represents where an equal proportion of male and female students answer correctly.

students. It is also in agreement with results from Chapter 3, which showed males achieving an overall higher FCI score. The improvement in student performance is shown through the general shift upwards (to the top right of plots presented in Figures 4.3 (a) and (b)) in overall correct responses. A narrowing of the gender gap can be seen through the clustering of data points closer to the line of unity in the post-instruction results. All but one data point still lay above the line of equality, illustrating that the gender performance gap remained measurable at the end of the semester. For some questions the gender gap was statistically significant in the post-test.

It was seen that the questions with the largest gender discrepancies are, for the most part, questions answered more poorly by the whole cohort. Table 4.1 shows the percentage of male and female students who answered a selection of FCI questions correctly in both the pre- and post-test. These questions, along with several other FCI questions, were chosen to demonstrate the different gender profiles seen on individual test items with respect to common misconceptions and are discussed in detail in the next section. Although the pre-test scores for each item differed amongst the three populations, with Manchester students performing the most highly in almost all cases, it can be seen that the level of performance and size of the gender gap is dependent on the individual test question.

Table 4.1: Percentage of 2011-12 male and female students correctly answering particular items on the pre-test and post-test. For Edinburgh N(male)=116 and N(female)=45. For Hull N(male)=40 and N(female)=6. For Manchester N(male)=198 and N(female)=60.

Item	Institution	Pre-test (%)		Post-test (%)	
		Males	Females	Males	Females
1	Edinburgh	85	78	93	89
	Hull	81	50	97	100
	Manchester	92	83	95	97
2	Edinburgh	65	40	81	62
	Hull	58	33	75	17
	Manchester	70	40	85	75
13	Edinburgh	54	31	91	80
	Hull	61	33	69	50
	Manchester	80	52	93	77
23	Edinburgh	78	40	85	64
	Hull	75	50	86	50
	Manchester	85	63	96	73
30	Edinburgh	54	42	85	80
	Hull	61	17	72	50
	Manchester	81	63	89	73

4.2 Multiple choice response profiles

Examining the incorrect multiple choice options chosen by students when answering questions can highlight difficulties faced with certain misconceptions. In the following section a representative sample of nine questions from the FCI will be discussed, along with analysis of the response profiles for each of these items. Of these nine questions, five contain schematic diagrams from which students are asked to consider the subsequent motion of an object after a force has been applied. The remaining four questions are descriptive and ask students to think about the existence and properties of the force(s) acting on an object. The following examples are by no means comprehensive but are representative of questions showing pre- or post-test gender gaps. Results from gender data at the Universities of Edinburgh, Hull and Manchester discussed in the previous section will be presented alongside indepth analysis of first year students at the University of Edinburgh.

Analysis of the multiple choice options from the combined data set of 2011-13 first year students from the University of Edinburgh has made clear which multiple choice options are more popular with male and female students. This in turn allows for the underlying misconceptions held by students as a whole to be identified and used for planning future teaching interventions. In each of the questions discussed we compare response profiles from a first year undergraduate cohort comprising 292 male students and 91 females students in the matched data set.

Cumulative frequency plots can be used to investigate the number of students who achieved each total test score and answered each multiple choice option correctly. The shape of the cumulative frequency plot can provide the instructor with an indication of the difficulty of the question or the prevalence of a certain misconception amongst students. Figures 4.4 (a)-(d) illustrate examples of possible cumulative frequency graphs. For example, the response profile illustrated in Figure 4.4 (a) demonstrates a scenario in which an equal proportion of students across all test scores answered this test option, resulting in a constant positive gradient on the cumulative frequency plot. Figure 4.4 (b) shows a straight horizontal line. This indicates that a small proportion of students with the lowest test scores chose this option, but that this option was not chosen by students with higher test scores. This may be indicative of a multiple choice option that demonstrates a clear misconception held by students with a lesser understanding of the topic. The response profile shown in Figure 4.4 (c) is slightly more complex. The increase in gradient at lower test scores is followed by a minimal increase for students in the middle performance quartiles. The gradient then increases for higher performing students. Finally, Figure 4.4 indicates a situation in which very few lower performing students chose this answer option. The sudden increase in gradient indicates that this option is nevertheless chosen by a high proportion of high performing students. Such a profile may indicate that this question may be used to help identify students who have a good understanding of the topic being tested.

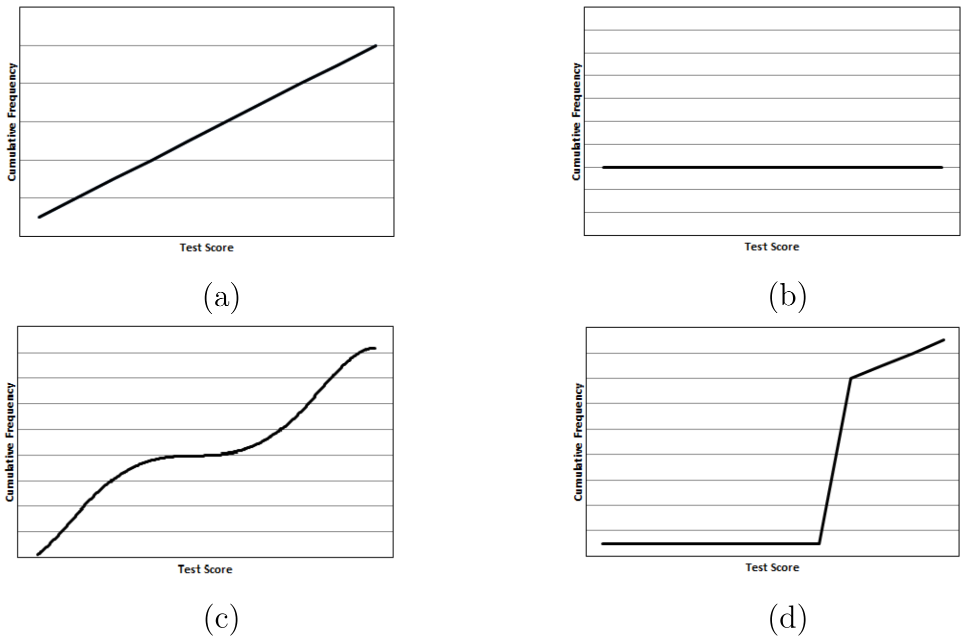


Figure 4.4: Examples of possible cumulative frequency graphs.

For each test item presented in this chapter the cumulative percentage of the male population and the cumulative percentage of the female population choosing each multiple choice option has been plotted against students' pre-test FCI score. For each answer option the number of male students was added as a cumulative percentage of the male population as you move up the FCI pre-test score axis. The same was done for the female population. The cumulative percentage for all possible multiple choice options should add to 100% of the cohort at the FCI score of 30 on the x-axis. Each FCI question contains five possible multiple choice answers. It is not necessarily the case that each of these five possible choices has been represented in the class' responses. Therefore, each graph represents only those choices which have been selected by participating students. A cumulative percentage of 50% of the cohort choosing the correct answer option was chosen as a baseline for comparison of male and female students.

Question 1

1. Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant. The time it takes the balls to reach the ground below will be

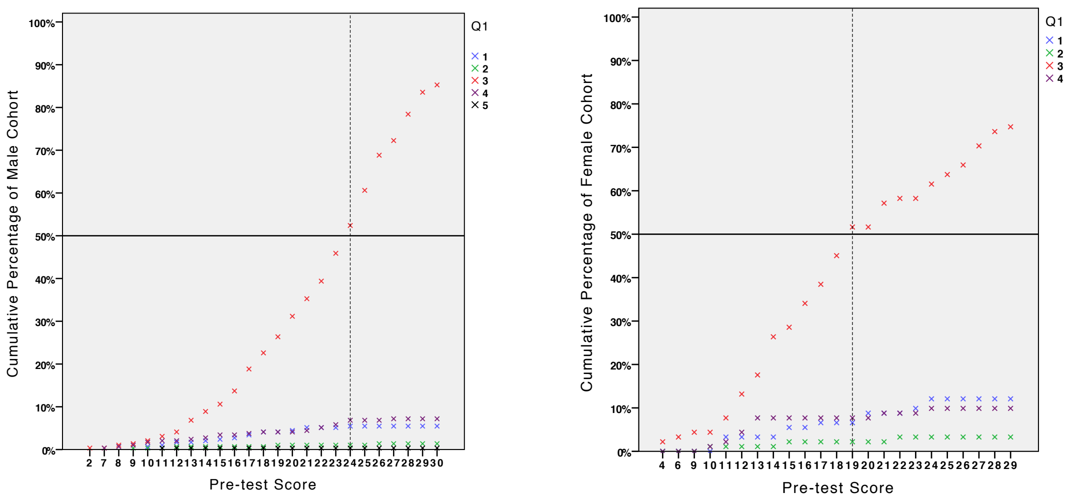
___ 1. about half as long for the heavier ball as for the lighter one.
___ 2. about half as long for the lighter ball as for the heavier one.
___ 3. about the same for both balls.
___ 4. considerably less for the heavier ball, but not necessarily half as long.
___ 5. considerably less for the lighter ball, but not necessarily half as long.

Figure 4.5: Question 1 from the FCI. See Appendix A for complete version of test [46].

The first two questions on the FCI are descriptive questions focusing on the motion of two balls of different weights after they have been dropped from a table at a time *t*. Both questions are descriptive with no accompanying diagrams or figures. In the first item, shown in Figure 4.5, two balls are dropped from the same height at the same time, with one ball being twice as heavy as the other. Respondents are prompted to choose from several options the relative time it will take the two balls to hit the ground. The correct response (option 3) states that the time will be the same for both masses; the principle that objects of different mass fall at the same rate.

This question was generally very well answered by all students. Looking first at results across the three universities in 2011-12, it was found that a slightly higher fraction of male students than female students answered the question correctly in the pre-test results (Table 4.1). Using a chi-squared test (correcting for a 2x2 table) it was found that this gender gap was not statistically significant at the 95% confidence level for students at any of the three universities. One must be aware of small number statistics at Hull, with only six female students in the matched data set, making it difficult to draw definitive conclusions from these results. In the post-test both genders showed a degree of improvement, with effectively no difference in the percentage of males or females answering correctly. In fact, in Hull and Manchester, females slightly outperformed males on this question at the end of the semester, again with the caveat of small number statistics at Hull.

Results from the analysis of Question 1 for the 2011-13 Edinburgh cohort are shown in Figures 4.6 (a) and (b). The graphs represent the cumulative percentage of each cohort selecting each multiple choice option as a function of their pre-test score. A comparison of the two answer profiles clearly shows that the majority of students was able to identify the correct response to the question. Despite this, there did exist a statistically significant difference between male and female students in the pre-test (p=0.014) for the combined year groups. In total 85% of male students and 75% of female students answered the question correctly at the start of the academic year. This difference was not found to be statistically significant in the post-test (p=0.796), with very high proportions of each cohort answering correctly. Although the other four possible answers are represented in the male results, and three in the female results, there did not appear to be one dominating multiple choice distractor, suggesting there may have been an element of guessing in the responses of students who answer incorrectly.



(a) (b)

Figure 4.6: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 1 of the FCI as a function of pre-test score. N(males)=292 and N(females)=91. The dotted lines indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

Despite a lower overall percentage of female students who answered this question correctly, it was seen that in order to reach a cumulative percentage of 50% of the total female population identifying the correct multiple choice option, all females with a pre-test score of 19 out of 30 or below had to be included. By comparison, this 50% baseline level for male students identifying the correct answer was reached by including all males with a pre-test score of 24 or lower. This suggests that a higher proportion of female students than male students in the lower and middle quartiles identified the correct answer on this question. Despite reaching this 50% baseline at a lower FCI score than males, the lower overall percentage of correct responses by the female cohort may reflect the small number of women in the top pre-test quartile. This may imply that the gender gap on this question originates from there being very few high performing women.

Question 2

2. The two metal balls of the previous problem roll off a horizontal table with the same speed. In this situation

___ 1. both balls hit the floor at approximately the same horizontal distance from the base of the table.

___ 2. the heavier ball hits the floor at about half the horizontal distance from the base of the table than does the lighter ball.

___ 3. the lighter ball hits the floor at about half the horizontal distance from the base of the table than does the heavier ball.

___ 4. the heavier ball hits the floor considerably closer to the base of the table than the lighter ball, but not necessarily at half the horizontal distance.

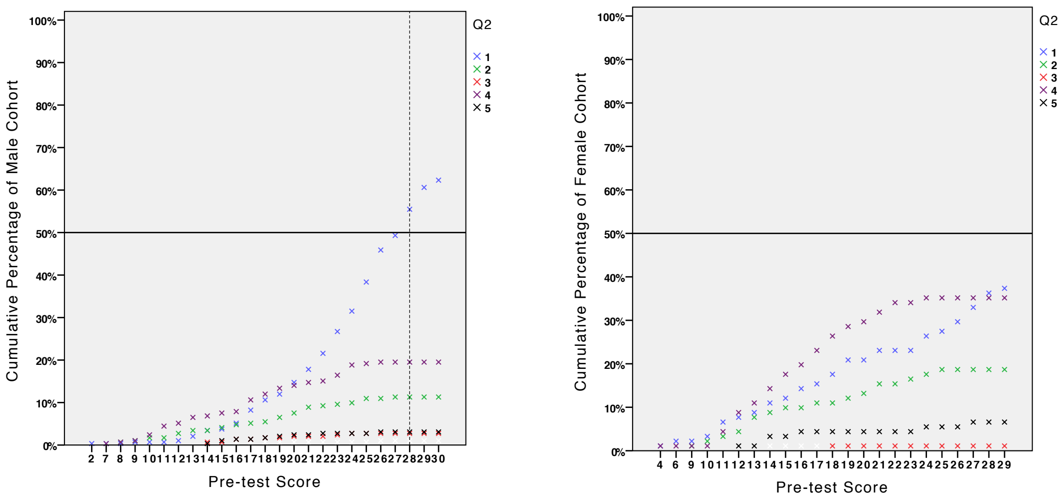
___ 5. the lighter ball hits the floor considerably closer to the base of the table than the heavier ball, but not necessarily at half the horizontal distance.

Figure 4.7: Question 2 from FCI. See Appendix A for complete version of test [46].

Question 2, shown in Figure 4.7, explores the same underlying physics principle as in Question 1. This question has been included to demonstrate a question in which students' answers are distributed across all possible multiple choice responses. It states that the two balls introduced in Question 1 roll off a horizontal table at equal speeds. Students are asked to choose from five statements describing the relative distance away from the table at which the balls hit the ground. Once again, the question is purely descriptive and contains no figures or diagrams in either the question or answer options. The correct answer, option 1, states that the two different masses will land at approximately the same horizontal distance from the edge of the table.

Although following on directly from the situation outlined in Question 1, Question 2 was answered less well by students at all three universities. Between 58-70% of male students in 2011-12 answered this correctly at the beginning of the course compared to only 33-40% of female students (Table 4.1). A statistically significant gender difference existed at both Edinburgh and Manchester universities pre-test (Edinburgh χ^2=7.099 and p=0.008, Manchester χ^2=16.780 and p<0.001) and this gender gap persisted post-instruction at Edinburgh (χ^2=5.268 and p=0.022). Interestingly, the number of females answering correctly at the University of Hull decreased in the post-test results, but again there is the caveat of small number statistics.

An independent t-test also showed a statistical difference between genders (p<0.001) for the 2011-13 first year undergraduate students at Edinburgh. This significant gender difference remained in the post-test results (p=0.007). The response profiles for male and female students' pre-test responses at Edinburgh are depicted in Figures 4.8 (a) and (b). When asked to determine the relative horizontal distance travelled by two balls of different masses, students showed a large degree of confusion, with both male and female answers distributed across all possible answer options.

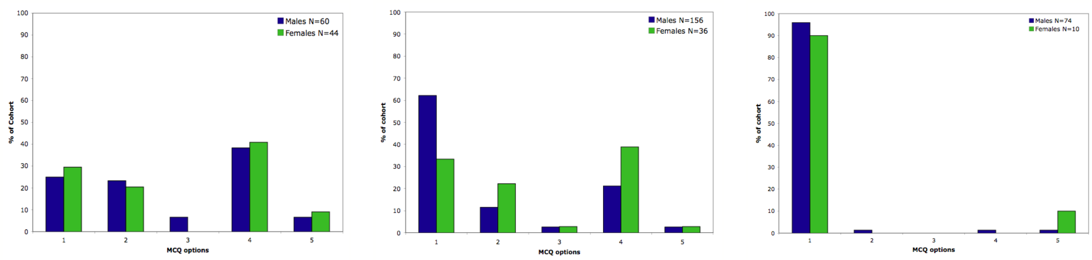


(a) (b)

Figure 4.8: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 2 of the FCI as a function of pre-test score. The dotted line indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

This question illustrated the contrast between male and female understanding of the effect of an object's mass on its projectile motion. Looking first at the results for male students, it was found that 50% of the whole male population choosing the correct answer had a test score less than or equal to 28 out of 30. Whilst this increased to a final 62% of the entire male population answering the question correctly, only 37% of females chose the correct answer in the first diet of the test.

The steep gradient of the slope corresponding to option 1 in Figure 4.8 (a) indicates that almost all of the males in the top performance quartile chose the correct option, with those in the lower quartiles showing a lesser degree of understanding. Responses from female students were distributed across all five of the answer options, with almost equal proportions of the cohort choosing options 1 and 4. Option 4 was the most popular incorrect response for both genders and states that "*the heavier ball hits the floor considerably closer to the base of the table than the lighter ball, but not necessarily at half the horizontal distance*". Here students incorrectly assumed that the horizontal distance travelled by an object is inversely proportional to its mass. For female students in the middle and lower performance quartiles this option was in fact the most popular response.



(a) (b) (c)

Figure 4.9: Percentage of male and female students selecting each multiple choice option in the (a) lowest quartile, (b) two middle quartiles and (c) top quartile of students for Question 2.

Graphs of the percentage of each gender cohort answering each option were created for each FCI pre-test quartile (Figures 4.9 (a) - (c)). For this analysis students in the two middle performance quartiles were combined into one data set because there existed very little difference between the response profiles of the two quartile populations. Figure 4.9 (a) shows that answers for students in the lowest performing quartile (pre-test score $\leq$16) were split across all five multiple choice options. This remained the case for the two middle quartiles (Figure 4.9 (b)), although the percentage of male students choosing option 1 was much greater than that of female students. Conversely, the percentage of females choosing option 4 was 18% higher than for males. This suggests that middle cohort females were more likely to carry the misconception of mass being inversely proportional to the distance travelled. Interestingly, almost all male and female students in the top performance quartile chose the correct response (Figure 4.9 (c)). This suggests that, perhaps unsurprisingly, the level of students' comprehension of this concept is related to their FCI pre-test score and that the main gender differences between answer choices for this question are by the middle performing students. Nevertheless, it is still necessary to target these misconceptions for the whole class since there is evidence of great confusion in the lowest quartile.

Question 5

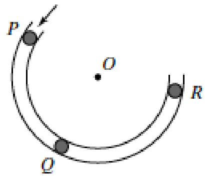
Question 5 of the assessment is the first of two questions discussing the motion of a ball shot through a frictionless u-shaped channel. Students are asked to refer to the accompanying diagram (Figure 4.10) and consider which of the listed forces act on the ball at the midpoint of its motion through the channel. The correct answer states that the ball experiences a downward force due to gravity as well as a centripetal force acting towards the centre of the circle of motion (option 2). Pre-test results showed that both male and female students struggled to identify the correct answer, particularly amongst the lowest performing students. This questions acts as an example of a test item in which less than half of the male and female populations were able to identify the correct answer. It also demonstrates that there existed differences between the chosen responses depending on gender.

Results from a comparison of gender performances showed no evidence of a statistical gender gap at any of the three participating universities in 2011-12 (Edinburgh χ^2=0.852 and p=0.356, Hull χ^2=0.004 and p=0.949 and at Manchester χ^2=3.883 and p=0.071). The pre-test scores for Question 5 were much higher for first year students at Manchester than at the other two institutions, with 75% of males and 62% of females answering correctly in the pre-test, compared to only 41% of males and 31% of females at Edinburgh and 38% of males and 50% of females at Hull.

Figures 4.11 (a) and (b) show the response profiles for each gender for 2011-13 students at Edinburgh. Students had a great deal of difficulty with this question,

Use the statement and figure below to answer the next two questions (5 and 6).

The accompanying figure shows a frictionless channel in the shape of a segment of a circle with its center at *O*. The channel has been anchored to a frictionless horizontal table top. You are looking down at the table. Forces exerted by the air are negligible. A ball is shot at high speed into the channel at *P* and exits at *R*.



5.Consider the following distinct forces:

A. a downward force of gravity.

B. a force exerted by the channel pointing from *Q* to *O*.

C. a force in the direction of motion.

D. a force pointing from *O* to *Q*.

Which of the above forces is (are) acting on the ball when it is within the frictionless channel at position *Q*?

___ 1. A only.

___ 2. A and B.

___ 3. A and C.

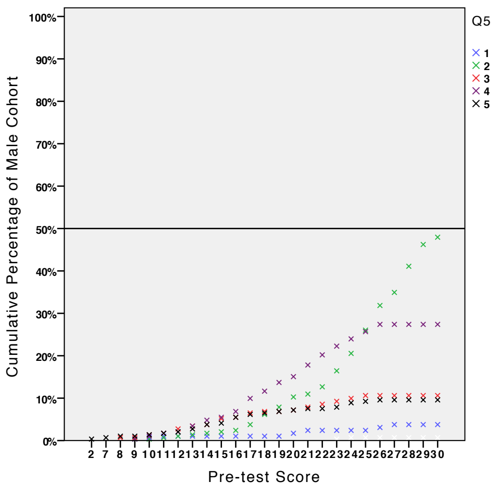
___ 4. A, B, and C.

___ 5. A, C, and D.

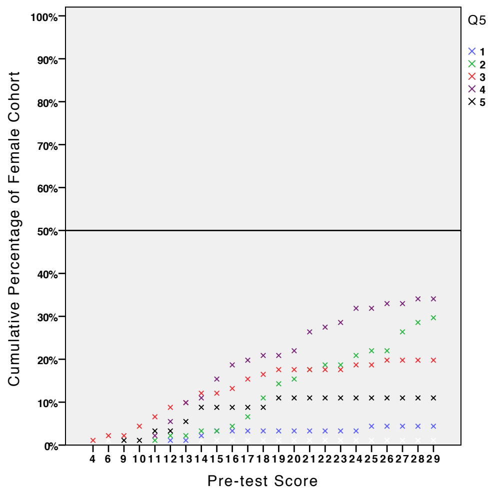
Figure 4.10: Question 5 from FCI. See Appendix A for complete version of test [46].

with both cohorts having relatively low pre-test scores. Their responses were not however dominated by one particular multiple choice distractor. Only 48% of the male population chose the correct answer in the pre-test. This value was much lower for the females, only 30% of whom answered correctly. The difference between the two cohorts was statistically significant in the pre-test (p=0.007).

Lower and middle performing male students showed a degree of uncertainty of the correct option. Almost all of the top performing males, those achieving a pre-test score of around 25 or higher, correctly identified the two forces acting on the object, therefore increasing the overall percentage of correct response for males. For the females, the story was slightly different. The most popular response for females was option 4; that there exists a downward force of gravity, a force exerted by the channel pointing towards the centre of the circle as well as a force



(a)



(b)

Figure 4.11: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 5 of the FCI as a function of pre-test score.

in the direction of motion. Female students with lower pre-test scores were more likely to believe that there existed a force in the direction of motion (option 3), and neglect the centripetal force, than the higher performing females. Those females with high pre-test scores struggled to realise that there was no force in the direction of motion. This is shown by the increase in cumulative percentage for females choosing answer options 2 and 4 for those students with the top pre-test scores, while the cumulative percentage of the other three options was stable. Both genders had high post-test scores and it was seen that there was no longer a statistical difference between the percentage of male and female students answering Question 5 correctly ($p=0.057$).

Questions 12 and 14

The next two questions to be discussed both refer to scenarios in which objects are falling from a height. In both cases students must decide what is the shape of the path taken by the falling object. Despite the similarities in these two questions, results collected indicated a large difference in the percentage of correct responses by students. In addition to this, Question 14 showed a large gender difference. Further analysis of the multiple choice response profiles for Question 12 and Question 14 may help highlight where students had the greatest difficulty and what were the most popular incorrect answers.

12. A ball is fired by a cannon from the top of a cliff as shown below. Which of the paths 1–5 would the cannon ball most closely follow?

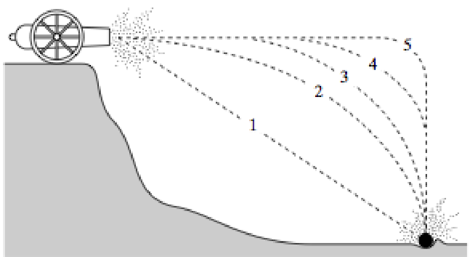
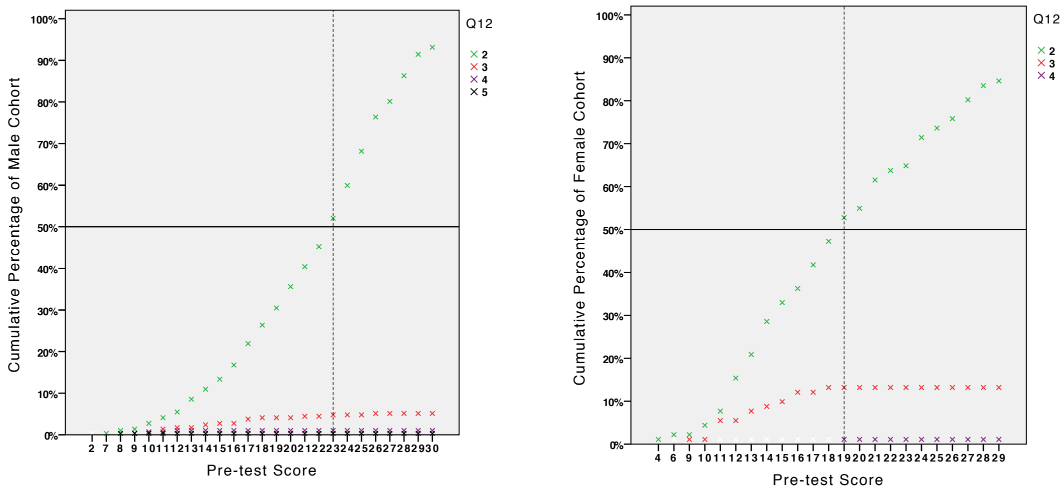


Figure 4.12: Question 12 from FCI. See Appendix A for complete version of test [46].

Question 12, shown in Figure 4.12, describes a scenario of a ball fired from a cannon at the top of a cliff. The accompanying diagram illustrates five possible paths taken by the cannon ball after it has been fired. Students are asked to identify which path they believe the cannon ball will follow, the correct solution being option 2.

This question was answered very well by both male and female students. Very high pre-test scores were observed for the Edinburgh, Hull and Manchester cohorts, with no statistical differences between gender at the 95% level. This was also the case for students at Edinburgh in 2011-13, where no statistically significant differences were found between male and female students in either the pre-test or post-test results (p=0.414 and p=0.647 respectively). Overall 93% of male students answered this item correctly in the pre-test compared to 85% of female students.

Although both cohorts showed high levels of conceptual understanding at the beginning of the semester, differences were found when comparing the 50% baseline measurement. Looking first at Figure 4.13 (a), it can be seen that a cumulative percentage of 50% of male students answering Question 12 correctly was reached by including all students with pre-test scores of 23 or lower. When the same analysis was completed for female students, it was found that a cumulative percentage of 50% of female students was reached by including all females with pre-test scores of 19 or lower (Figure 4.13 (b)). Once again this highlights the fact that there was a higher proportion of females in the lower performing



(a) (b)

Figure 4.13: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 12 of the FCI as a function of pre-test score. The dotted lines indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

pre-test quartiles. These lower performing females were more likely to hold a misconception about the parabolic motion of the cannon ball. The most common distractor was option 3, which depicts the cannon ball continuing for a short time in the horizontal direction before feeling the downward effect of gravity. While 13% of female students chose this option, only 5% of male students chose it.

Question 14 is closely related to Question 12 of the FCI. Similarly to the scenario of the cannon ball fired from the top of the cliff, this question introduces a bowling ball falling out of an airplane which is flying in a horizontal direction, as shown in Figure 4.14. Once again students are presented with five possible paths taken by the bowling ball, as viewed by a spectator on the ground, and asked to identify the path most closely followed by the ball. The correct answer is option 4.

Surprisingly, the pre-test response rates differed considerably to those of Question 12, in which the vast majority of students had little difficulty in discerning the correct path. Edinburgh and Manchester showed large gender gaps between students' pre-test scores. At Edinburgh, 73% of males answered correctly compared to only 47% of females (χ^2=9.058, p=0.003). At Manchester scores were slightly higher, with 79% of males answering correctly compared to 65% of females (χ^2=4.399, p=0.036). At Hull the proportions of males and

14. A bowling ball accidentally falls out of the cargo bay of an airliner as it flies along in a horizontal direction.

As observed by a person standing on the ground and viewing the plane as in the figure below, which of the paths 1–5 would the bowling ball most closely follow after leaving the airplane?

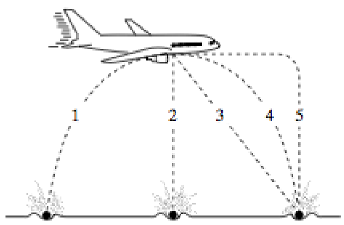
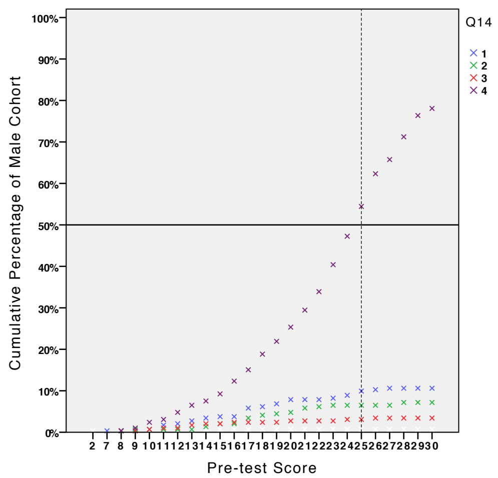


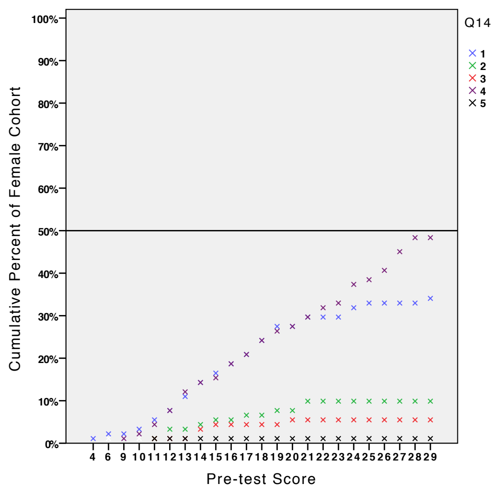
Figure 4.14: Question 14 from FCI. See Appendix A for complete version of test [46].

females who chose the correct path were very similar; 64% of males and 67% of females. This gender difference was not significant (χ^2=0.017, p=0.895).

At Edinburgh between 2011-13, unlike for Question 12, Question 14 illustrated a statistical gap between male and female students' performance, both in the pre-test (p=0.004) and the post-test (p<0.001). For this question a total of 78% of the male cohort correctly chose option 4 in the pre-test, compared to only 48% of females (Figures 4.15 (a) and (b)). The 22% of males who answered incorrectly was split between options 1, 2 and 3. Looking at the female responses it was seen that, for students with pre-test scores below 22, choosing option 1, in which the bowling ball is seen to travel in the negative horizontal direction, was equally popular to choosing the correct path. The high number of responses for option 1 suggests that respondents were unaware that the bowling ball's horizontal motion will remain unaffected after it has been released from the airplane, and it will continue to travel with the same horizontal velocity. The force of gravity acts in a direction perpendicular to that of the ball's horizontal motion and therefore has no influence on its horizontal motion. Perhaps surprisingly there were very few students who chose option 2, in which the ball falls vertically downwards. It could be argued that this answer option depicts the fact that the object will land directly below the plane as they both continue to have the same horizontal velocity. This question provides an opportunity for further investigating students' misconceptions. Results from qualitative interviews with students, in which students' interpretation of this diagram are explored, will be discussed in more detail in Chapter 5.



(a)



(b)

Figure 4.15: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 14 of the FCI as a function of pre-test score. The dotted line indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

Question 13

Question 13 is a descriptive question (Figure 4.16) in which students consider which force(s), along with their magnitudes and directions, act on a ball after it is thrown vertically upwards and released from someone's hand. Once the ball has left the hand, if air resistance is neglected, the only force acting on the object is the downward force of gravity, as stated by option 4. Alternative multiple choice answers reflect a common misconception amongst students in which the force of the 'throw' given to the ball by the thrower persists (either as a constant or steadily decreasing force) after the object has left the thrower's hand.

Results from Table 4.1 showed that, at all three institutions, there existed a clear gender discrepancy between the number of male students and the number of females students correctly identifying that only the force of gravity acts on the ball. Statistically significant differences between male and female pre-test scores were found at Edinburgh and Manchester (Edinburgh χ^2=5.626 and p=0.018,

13. A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand but before it touches the ground, and assume that forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are)

___ 1. a downward force of gravity along with a steadily decreasing upward force.

___ 2. a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the ball gets closer to Earth.

___ 3. an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only an almost constant downward force of gravity.

___ 4. an almost constant downward force of gravity only.

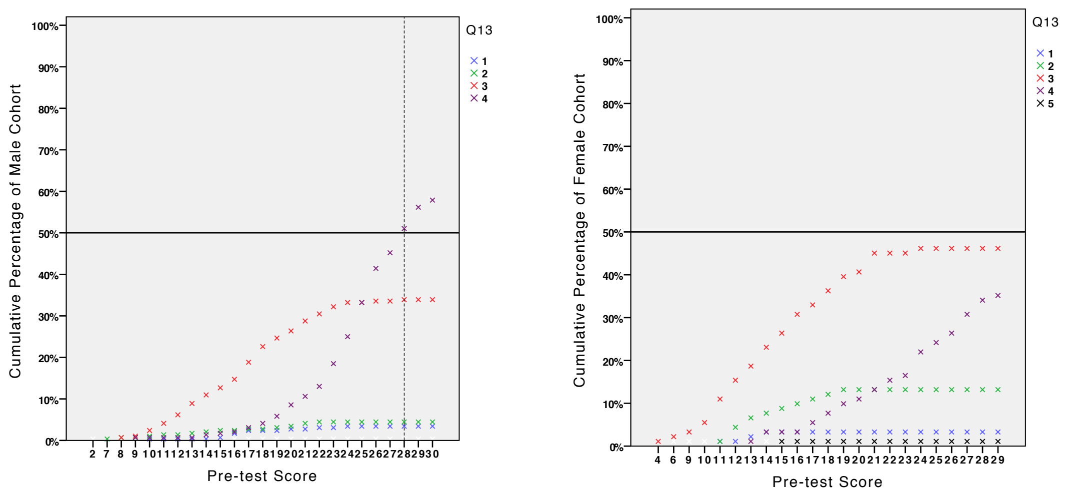
___ 5. none of the above. The ball falls back to ground because of its natural tendency to rest on the surface of the Earth.

Figure 4.16: Question 13 from FCI. See Appendix A for complete version of test [46].

Manchester χ^2=16.998 and p<0.001) in 2011-12. Despite a great improvement in both male and female scores in the post-test, there persists a noticeable gender gap. This gap was significant for male and female students studying at Manchester (χ^2=12.085 and p=0.001).

Graphs of male and female response profiles (Figures 4.17 (a) and (b)) showed that only 35% of the total female population at Edinburgh in 2011-13 responded correctly in the pre-test. This was higher for males (58%). An independent t-test found this difference to be statistically significant (p=0.003), and this significant gender gap persisted in the post-test results (p=0.007). A steep increase in the gradient of the cumulative percentage at high pre-test scores, suggesting that both male and female students in the top performance quartile were most likely to choose the correct answer.

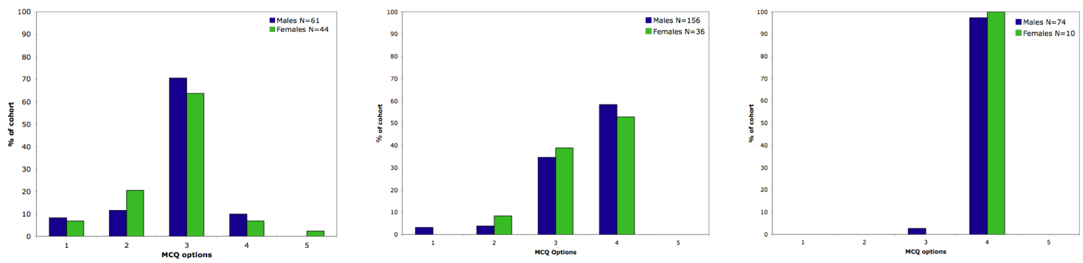
Figures 4.17 (a) and (b) demonstrated that there were several underlying misconceptions held by students. In particular multiple choice option 3 was the most popular choice in the lower and middle quartiles for both males and females. In fact, the majority of female students chose option 3. This option states that there exists "*an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only an almost constant force of gravity*". By choosing this answer students demonstrated that they are holding on to the misconception that there exists a lingering effect of the upward force given to the ball by the thrower,



(a) (b)

Figure 4.17: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 13 of the FCI as a function of pre-test score. The dotted line indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

even after it has left their hand. This is often referred to as the misconception that motion implies force; the assumption that a force must always be present to sustain motion in a defined direction. The dominance of option 3 persisted for lower and middle performing students, but the gradient of the cumulative percentage of responses for this option did not increase when top performing students were included in the analysis, implying that students with higher pre-test scores have overcome this misconception.



(a) (b) (c)

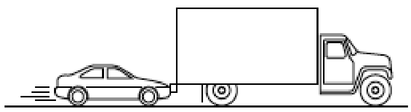
Figure 4.18: Percentage of male and female students selecting each multiple choice option in the (a) lowest quartile, (b) two middle quartiles and (c) top quartile of students for Question 13.

The prevalence of this misconception is reflected in graphs of the percentage of males and females in each quartile choosing each option (Figures 4.18 (a) - (c)).

Each of the five possible options was represented in responses by lowest quartile students, with the incorrect option 3 the most popular for both genders. In the middle quartiles students' responses were effectively split between two possible answers (options 3 and 4). Finally, looking at the top quartile, approximately 90% of both cohorts correctly identified gravity as the only force acting on the ball. Once again this suggests that students' FCI pre-test scores can be directly linked to the degree to which they hold onto this preconception of force and motion.

Question 15

A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the figure below.



15. While the car, still pushing the truck, is speeding up to get up to cruising speed,

___ 1. the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.

___ 2. the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car.

___ 3. the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.

___ 4. the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.

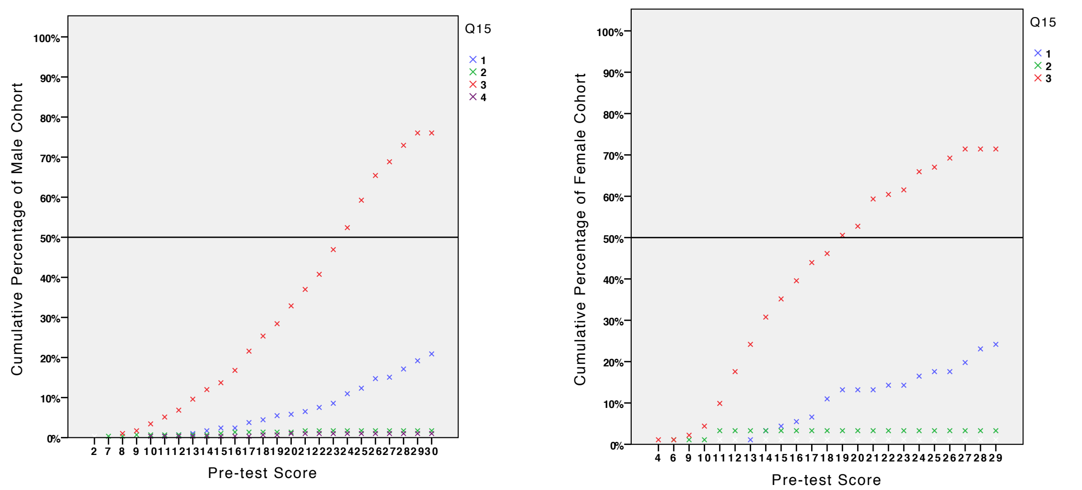
___ 5. neither the car nor the truck exerts any force on the other. The truck is pushed forward simply because it is in the way of the car.

Figure 4.19: Question 15 from FCI. See Appendix A for complete version of test [46].

Question 15 is another example of a question in which responses were split between the correct answer (option 1) and one dominating misconception. This question focuses on the concept of Newton's Third Law, the principle that each force has an equal and opposite reaction force.

Students at all three universities showed very low levels of understanding of Newton's Third Law. At Edinburgh, equal proportions of the male and female populations chose the correct answer (22%), whilst at Manchester 34% of males and 30% of females answered correctly. At Hull, 17% of males correctly chose option 1 whilst all six of the females answered incorrectly. In each case these gender differences were not significant at the 95% percentile as calculated by a chi-squared test (Edinburgh χ^2=0.009 and p=0.926, Hull χ^2=1.167 and p=0.280 and at Manchester χ^2=0.334 and p=0.563).

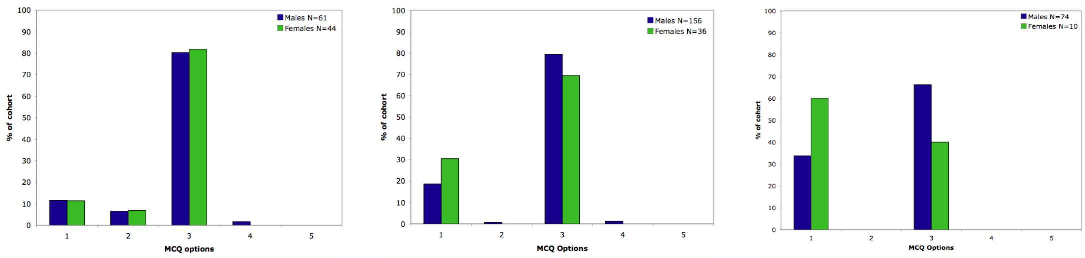
Figures 4.20 (a) and (b) show the response profile graphs for male and female students at Edinburgh between 2011-13. Both cohorts had remarkably low percentages of students answering correctly in the pre-test; only 21% of males and 24% of females. There was no statistical difference between the genders either prior to or post instruction (p=0.231 and p=0.717 in the pre- and post-tests respectively). The vast majority of participants chose the incorrect statement that "*the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car*", incorrectly assuming that since the car and truck are speeding up the car must exert the greater force to push the truck. Students' understanding of this fundamental principle is further investigated in a qualitative study discussed in Chapter 5.



(a) (b)

Figure 4.20: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 15 of the FCI as a function of pre-test score.

The quartile response graphs, shown in Figures 4.21 (a) - (c), are of particular interest. Over 80% of students in the lowest quartile held the misconception that movement or acceleration in a defined direction implies that one object exerts a greater force than the other. Only approximately 10% of students were aware of the consequences of Newton's Third Law on this situation. When the responses from the middle quartile were examined, the proportion of students who chose the correct answer increased, but still remained remarkably low. Interestingly, the percentage of female students answering correctly was higher than for males. The top quartile showed opinions to be polarised between two responses. It is clear that females are more likely to choose option 1 over males, as was reflected in the final pre-test percentage of correct answers. The slightly higher percentage of female students answering correctly in both the middle two quartiles and the top quartile led to the slightly higher proportion of females answering correctly overall. There was a smaller proportion of females in the top performance quartile which resulted in a small overall percentage gender gap.



(a) (b) (c)

Figure 4.21: Percentage of male and female students selecting each multiple choice option in the (a) lowest quartile, (b) two middle quartiles and (c) top quartile of students for Question 15.

The fact that Newton's Third Law is a physics principle with which many students have problems is well documented in the literature [136, 137, 138]. Newton's Third Law deals with the concept that two forces arise purely from interactions. A study by Brown noted that high school students entered class holding several misconceptions about this principle and that these preconceptions remained at the end of the teaching semester [136]. They commented that these may result from "*students' general naive view of force as a property of single objects rather than as a relation between objects*".

Question 23

One FCI question which showed contrasting gender response profiles in 2011-13 was Question 23. Although a high proportion of male students identified the correct response, female students showed a greater degree of confusion,

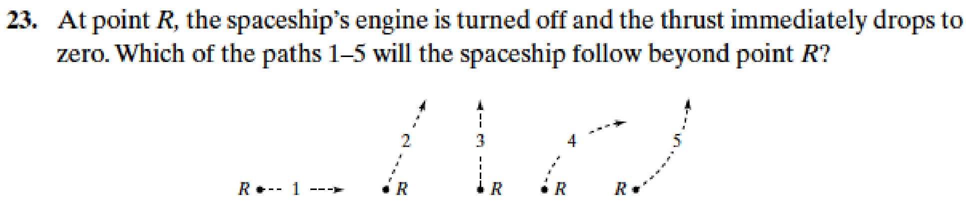


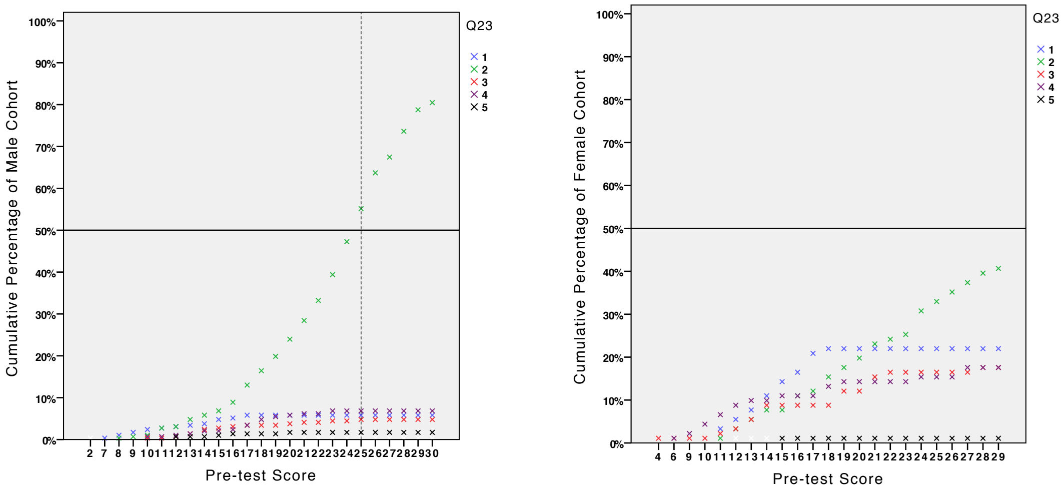
Figure 4.22: Question 23 from FCI. See Appendix A for complete version of test [46].

particularly in the lower performing quartiles, the results of which are discussed below.

Question 23 forms the third part of a set of four consecutive questions referring to a series of schematic diagrams, and is representative of several items on the assessment instrument. In this series of questions participants are asked to determine the effect of introducing or removing a force on an object in the absence of friction. These sorts of questions, combining uniform motion in one direction with an accelerating force applied in one perpendicular, tend to cause students a significant challenge. Students are presented with a situation in which a spaceship is drifting horizontally in space with no outside forces acting on it. In the previous questions a force is applied to the spaceship at right angles to its current path by turning on its engines. Question 23 asks students to identify the path of the spaceship, this time after the engine has been switched off and the thrust eliminated, with option 2 being the correct response.

Question 23 in particular exhibits large gender differences in pre-test scores at the Universities of Edinburgh, Hull and Manchester, as shown in Table 4.1. Between 75% and 85% of male students across the three universities answered correctly at the beginning of the semester compared to between 40% and 63% of female students. A chi-squared test found that the initial significant gender gap at the start of the 2011-12 semester in Edinburgh and Manchester (Edinburgh χ^2=19.075 and $p<0.001$, Manchester χ^2=12.580 and $p<0.001$) remained after a semester of teaching, with relatively small improvements made in male and female post-test scores (Edinburgh χ^2=7.456 and $p=0.006$, Manchester χ^2=25.323 and $p<0.001$).

Considering the combined 2011-13 population at Edinburgh separately, it was found that this item in particular exhibited a large, and statistically significant gender gap pre-instruction ($p<0.001$); 81% of males answered this correctly



(a) (b)

Figure 4.23: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 23 of the FCI as a function of pre-test score. The dotted line indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

compared to 41% of females (Figures 4.23 (a) and (b)). Although there was evidence of improvements by both male and female cohorts post-instruction, this gender gap remained significant post-test with a p value of <0.001. The response profiles for male students across all quartiles is dominated by the correct answer, with each of the other options being chosen by less than 8% of the cohort. Conversely, the female response profiles showed a larger degree of confusion across all pre-test scores. A higher proportion of female students in the lower and middle pre-test quartiles chose option 1 than the correct answer (option 2). Overall 22% of females chose option 1 compared to only 6% of males. This may suggest that females may hold on to the misconception that all motion requires an acting force.

Question 30

As was the case in Question 15, the 2011-13 Question 30 response profiles indicate that students' responses are dominated by one common misconception, which is predominately chosen by students with lower overall FCI pre-test scores.

In Question 30 of the FCI a tennis player hits a ball with a racket in the presence of a high wind (Figure 4.24). Students are asked to decide what force(s) are acting on the tennis ball during its motion. After the ball has left the racket, the only forces acting on it are the downward force of gravity and the force exerted by the air, as stated in option 3. This question tests students' understanding of motion after the removal of an applied force.

30. Despite a very strong wind, a tennis player manages to hit a tennis ball with her racquet so that the ball passes over the net and lands in her opponent's court.

Consider the following forces:

A. a downward force of gravity.

B. a force by the "hit."

C. a force exerted by the air.

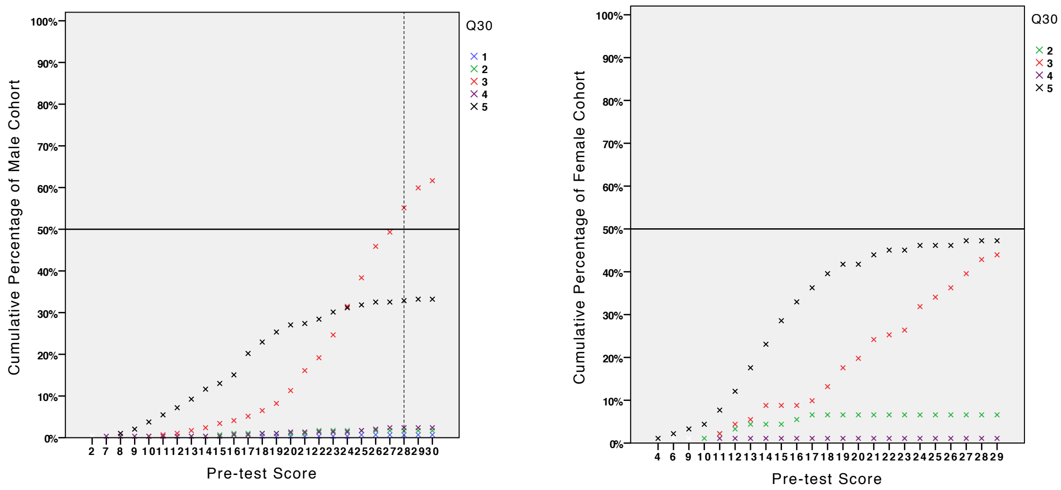
Which of the above forces is (are) acting on the tennis ball after it has left contact with the racquet and before it touches the ground?

___ 1. A only
___ 2. A and B
___ 3. A and C
___ 4. B and C
___ 5. A, B, and C

Figure 4.24: Question 30 from FCI. See Appendix A for complete version of test [46].

Results from the three university cohorts showed large variations in 2011-12 pre-test scores. At the University of Edinburgh male students outperformed female students, but not significantly (p=0.230). At the beginning of the semester, 54% of males students identified the correct answer compared to 42% of females. The average pre-test result was slightly higher for males at Hull than at Edinburgh. Although only one of the six females was correct, 61% of the male population answered correctly. Despite having a much higher percentage of correct pre-test responses to this question (81%), relatively small improvement was made by male students at the University of Manchester compared to male students at the other institutions, in particular Edinburgh. Interestingly, there did exist a significant gender gap between male and female students at Manchester both prior to instruction (χ^2=6.930 and p=0.008) and after a semester of teaching (χ^2=6.439 and p=0.011).

Intriguing gender differences were seen in the 2011-13 population at Edinburgh. In total 62% of the male population chose the correct response. Only 44% of the female cohort chose the correct option. Question 30 showed clear evidence for the existence of a misconception by students. Figures 4.25 (a) and (b) demonstrate that both genders believed there to be a persisting effect resulting



(a) (b)

Figure 4.25: Cumulative percentage of (a) male and (b) female students selecting each multiple choice option for Question 30 of the FCI as a function of pre-test score. The dotted line indicates the FCI pre-test score associated with a cumulative percentage of 50% of the male and female cohorts answering correctly.

from the force supplied by the racket to the ball, a similar effect to that seen in Question 13 when a ball was thrown vertically upwards from a boy's hand. This misconception was particularly prevalent in responses by students with pre-test scores below 20. Approximately 7% of female students did not appreciate the fact that there would exist a force exerted on the ball from the strong wind. Although 62% of male students responded correctly overall, only 44% of females were able to identify the correct response.

4.2.1 Summary

These questions (Table 4.2), together with the proportion of male and female students answering other test questions correctly, and the resulting distributions of answer choices, illustrate a complex picture of the behaviour of male and female students. The position of the 50% cumulative frequency baseline was determined for both the male and female populations. This gave an indication of the comparative understanding and performance of students depending on their FCI pre-test scores. As can be seen in Table 4.2, in many cases less than 50% of the female population successfully answered the questions discussed in this chapter. This provides a good indication of questions in which males significantly outperformed females.

Table 4.2: Percentage of male and female students answering individual FCI questions correctly and the pre-test scores corresponding to 50% cumulative percentage of the cohort answering correctly.

Question Number	Male Students		Female Students	
	% Correct	50% baseline pre-test score	% Correct	50% baseline pre-test score
1	85	24	75	19
2	62	28	37	-
5	48	-	30	-
12	93	23	85	19
13	58	28	35	-
14	78	25	48	-
15	21	-	24	-
23	81	25	41	-
30	62	28	44	-

The selected questions have highlighted the existence of common misconceptions amongst students. Many students were seen to believe that a force continues to act on an object even after it is no longer in contact with the agent supplying the force. Clement noted this preconception amongst engineering students taking a compulsory introductory mechanics course [131]. When asked to draw the direction of the force acting on a coin tossed upwards, students had difficulty reconciling that the object continues to move in a direction opposite to the force acting on it. This was the case for Question 13 in which a boy throws a steel ball vertically into the air. A high proportion of students at Edinburgh believed that the force applied by the boy steadily decreased during the upward motion. Interestingly, this was not the case for female students for Question 23. Over 20% of the cohort believed that after its engine was turned off the spaceship's original motion would be unchanged, and it would continue to move in a horizontal direction. This question also highlighted the fact that the prevalence of certain misconceptions sometimes differed for male and female students. Some misconceptions, such as the belief that dropped objects lose their forward motion and have no impetus, as seen in Question 14, were not held equally by both genders, with lower performing female students often showing a larger degree of confusion.

Overall, results suggest that female students, particularly those scoring in the lower FCI pre-test quartiles, are more likely to be affected by preconceived misconceptions than their male counterparts. There is also an indication that, in the majority of cases, the answer options chosen by students are not randomly distributed across multiple choice options but are instead rooted in common errors in students' conceptual understanding. The origins of these misconceptions and students' interpretation of the test items will be discussed in more detail in Chapter 5.

4.3 Student transitions between pre- and post-test

In addition to noting whether students answered individual test items correctly or incorrectly, the transitions between these two options can be examined. Results were combined for the 2011-12 and 2012-13 data sets from the University of Edinburgh, both of which used the same version of the FCI. The student transitions for all of the FCI questions from these two year groups were binned into four categories; right-to-right, right-to-wrong, wrong-to-right and wrong-to-wrong. A 'right-to-right' transition indicates that a student chose the correct answer in the pre-test and also selected the correct answer in the post-test. Similarly, a 'right-to-wrong' transition indicates that the correct answer was chosen in the FCI pre-test but was subsequently changed to an incorrect response in the post-test. It is possible the 'wrong-to-wrong' transition may include either students choosing an incorrect response in the pre-test and the same incorrect response in the post-test, or students changing their answer between two incorrect responses. For the purpose of this analysis these two options have been collapsed into the same category. For each of these categories the total number of such transitions undertaken by students was calculated, and subsequently split by gender.

Figure 4.26 shows the distributions of male and female students' transitions. A chi-squared test of their distributions in these four categories showed a statistically significant difference between genders ($p<0.001$). We can see that overall a large proportion of students answered correctly in both the pre-test and

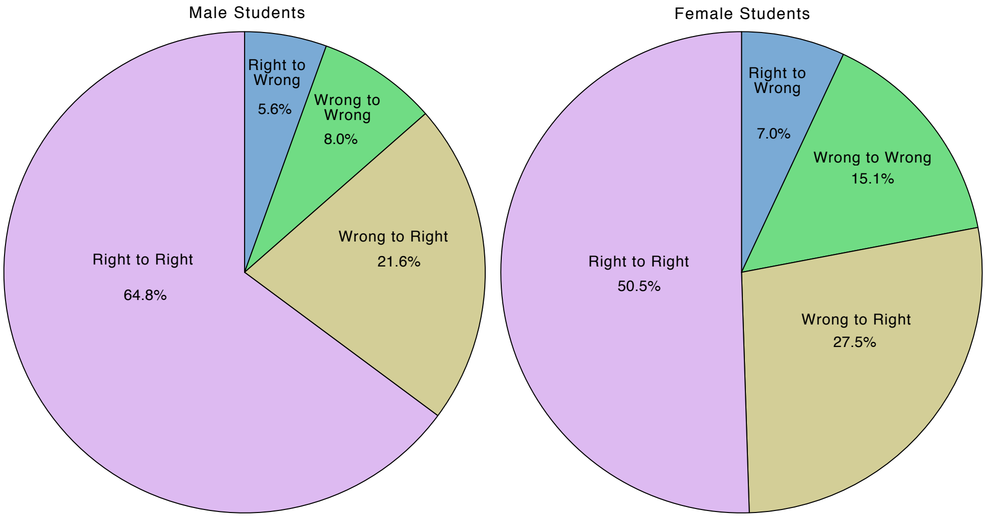


Figure 4.26: The percentage of male and female students making transitions between correct and incorrect multiple choice answers for the combined 2011-13 FCI data set.

post-test. There are several points of interest when comparing the distribution of male and female responses. The percentage of 'right-to-right' answers for male students (64.8%) was much higher than for female students (50.5%), consistent with the fact that male students were more likely to have a higher score in the pre-test. It can be said that the 'right-to-right' and 'wrong-to-wrong' responses have no effect on the measured change, or gain, between pre-test and post-test scores for cohorts. The probability of a male student maintaining an initially correct answer was 92%, compared to female students for whom the probability was 88%. Similarly, the probability of a student who initially answered incorrectly also answering incorrectly post-instruction was 27% for males compared to 35% for females.

In order for a numerical gain in cohort scores to be witnessed between the two administrations of the test, the percentage of 'wrong-to-right' transitions must be larger than the percentage of 'right-to-wrong' transitions. This was the case for both male and female students. We do however see that there was a larger percentage difference between these two categories, in favour of wrong-to-right, for the female students at Edinburgh than for males. This reflects the previously presented data from Chapter 3 which showed that, on average, the female population had higher normalised gains on the FCI.

Although the 'right-to-wrong' category showed low percentages of transitions for male and female students, 5.6% and 7.0% respectively, their effect is not negligible. This suggests that some students had a negative gain on some FCI questions. They initially answered the question correctly but, after a semester of teaching, changed this answer to an incorrect response.

A study by Lasry *et al* tested the stability of student responses on the FCI by similarly categorising answer transitions [139]. They commented that a 'right-to-wrong' transition could be indicative of either a false-positive (in which students answer correctly but have an error in their conceptual understanding) on the pre-test or a false-negative (in which students' correct reasoning was not reflected in their vote) on the post-test. Ideally 'right-to-wrong' transitions should be minimised in order to achieve the highest possible learning outcome in a course. Results presented in this chapter are the percentage transitions for all 30 FCI combined. Results for individual questions naturally showed variation dependent on the item and the concept tested.

4.4 Chapter discussion and summary

In this chapter the extent of the existing gender gap witnessed in the conceptual understanding of first year undergraduate physics students at University of Edinburgh has been explored, firstly through a comparison of the percentage of correct male and female responses for each FCI question, and secondly through comparing response profiles for a selection of test items. Both male and female students showed a high level of improvement between pre- and post-test FCI results. It was queried whether the observed gender disparity in conceptual understanding derived from males outperforming females across the whole assessment or whether they significantly outperformed on specific questions. Analysis of data from both the 2006-10 and 2011-13 cohorts concluded that males outperformed females on each question in pre-test results. This gender difference was significant in many cases, suggesting that females had a lesser understanding of specific concepts, and underperformed across the assessment as a whole.

Many similarities were noted when studying data collected from other UK universities. In addition to the overall persisting gender gap noted in Chapter 3, combining data from the universities of Edinburgh, Hull and Manchester indicated that the vast majority of FCI questions had a higher proportion of males answering correctly compared to females. Despite differences in the pre-test scores amongst the three institutions, it was noted that the level of performance varied considerably depending on the test item. Several questions showed statistically significant gender gaps across the different populations, indicating that this result was not unique to the University of Edinburgh nor dependent on the delivery method of the instrument. A comparison with the data from the University of Minnesota suggested that many of the same items have a gender difference for students from both the UK and USA [62]. We can extract only the fraction of correct answers from their paper, so are unable to judge whether the same answer distractors have been chosen in both cases. If considering potential cultural differences, results from a study of 10th grade Turkish high school pupils showed very little difference between male and female students' conceptual understanding as measured by the FCI [140]. These results were collected from students at a very early stage in their science education and the low levels of performance on the test overall do not allow for direct comparisons to be made, but may possibly suggest that cultural factors play a role in gender performance.

In terms of misconceptions of force and motion, students in the first year introductory course at the University of Edinburgh demonstrated a complex picture, with many questions showing contrasting response profiles for different genders. For example, one of the test questions discussed highlighted the difficulty in students' ability to understand that an object's mass has no effect on the horizontal distance travelled in projectile motion. Female participants were considerably more likely to answer incorrectly and, perhaps unsurprisingly, students with the lowest pre-test scores showed the highest levels of confusion, with almost all answer choices represented in student responses. Another prevalent misconception was that all motion must imply the presence of a force. In Question 13, when a ball is thrown upwards from a boy's hand, only 35% of females and 58% of males concluded that gravity alone acted on the ball during its flight. The most common misconception was that there existed a steadily decreasing upward force on the ball until it reached its highest point. Some physics concepts, such as Newton's Third Law, remain challenging for all first year students, suggesting more emphasis on teaching this principle may be required. The difficultly students have with applying the Newton's Third Law principle is well documented in published literature [136, 137, 138]. This analysis aided in the choosing of test items to be used in qualitative interviews with students enrolled on first year physics courses. These interviews aimed to develop a greater understanding of the origins of these misconceptions and will be discussed in detail in Chapter 5.

The context of the questions in the FCI (for example rockets, cars and cannons) raises the question whether the FCI instrument is partially to blame for these difficulties. There has been at least one attempt to make a less male-stereotyped version of the FCI [75], though results using this were largely inconclusive due to a low overall attainment, both pre- and post-instruction, on the refined instrument, obscuring any potential real effect. Moreover, we see substantial gender differences in our results on questions that do not have gender-stereotyped contexts.

Literature has shown that there may exist a gender bias on multiple-choice assessments themselves. A study conducted by Bolger and Kellaghan indicated that male school students performed significantly better than females on multiple-choice assessments compared to open response tests, both in languages and mathematics [71]. Some studies of open response assessments in both STEM subjects and the humanities have shown that women usually performed better than males [141, 142]. Possible explanations for the female bias towards open response tests have been the high verbal and written skills of females or gender differences in risk-taking tendencies [142]. It has also been suggested that differences in student performance on multiple-choice and open response tests may be an effect of the restrictions placed on the content that can be examined using a multiple choice format and that multiple-choice tests require a lower cognitive demand [143].

Results from analysis of students' transitions between pre- and post-testing offer reassurance of overall learning gains. Averaged across the whole assessment, a high proportion of students selected the correct response to a question in both the pre- and post-test. There did exist a statistical difference between the distribution of male and female students across the four possible transition groups. Female students had a higher percentage of 'wrong-to-right' transitions, contributing to the overall higher measured normalised gain on the test. One concern is the proportion of students continuing to have 'wrong-to-wrong' transitions. They accounted for 15% of the female transitions and 8% of the male transitions. Targeting these students and identifying their misconceptions could improve overall course results. Of even more concern is the existence of students residing in the 'right-to-wrong' category. These students may be those who were guessing, either in the pre- or post-test, or those who have increased misconceptions post-test. It is particularly important to find out if the latter exist.

The analysis included in this chapter opens up several avenues for further work. When new concepts are introduced during teaching it may be necessary to overcome preconceived beliefs that students have built up from prior learning. An awareness of these issues may lead to the conclusion that more emphasis be placed on addressing these fundamental misconceptions of physics principles explicitly at the point of entry to introductory physics courses. This quantitative analysis has enabled key test questions and misconceptions to be identified for further instruction. In order to determine a course of action for targeting such misconceptions, it is necessary to fully comprehend the source of students' preconceptions. There exists a need to characterise people's reasoning about certain physics concepts. In doing so further understanding may be achieved about which physics concepts may be conflicting with students' 'common-sense' beliefs. Students' reasoning and interpretation of FCI questions will be investigated in Chapter 5 of this thesis. Specific misconceptions highlighted in response profiles of the questions presented in this chapter will be explored through qualitative interviews.

Chapter 5

Qualitative Analysis of Misconceptions

A misconception can be defined as a belief that contradicts scientific fact or reasoning which students bring to their learning experiences. Students' misconceptions can originate from a variety of sources including the individual's personal experiences or prior teaching [132]. Misconceptions in physics, particularly in classical mechanics, are often strongly related to everyday 'real world' experiences. Halloun and Hestenes stated that

"Common sense beliefs about motion are generally incompatible with Newtonian theory. Consequently, there is a tendency for students to systematically misinterpret material in introductory physics courses. Common sense beliefs are very stable, and conventional physics instruction does little to change them." [52]

In topics such as electromagnetism, which often draw very little relation to personal experiences, the existence of such misconceptions may be explained as an incomplete knowledge or understanding of the topic, rather than drawing on the 'real world' experiences that can influence understanding of Newtonian mechanics. Results from a test of students' understanding of electromagnetism were presented in Chapter 3 and showed no evidence of a gender gap.

In the previous chapter results from the answer profiles of a representative sample of Force Concept Inventory (FCI) questions were presented. The clear trends in students' answers, and the distractor options chosen in the multiple choice test, indicate that in most cases these incorrect answers are not randomly distributed but are rooted in common errors in students' conceptual understanding, rather than the result of students guessing. Often these underlying misconceptions can go unnoticed due to a student's ability to use quantitative techniques and formulae to solve a problem [131]. If these inconsistencies between physics concepts and students' intuition can be identified, actions can be taken to change these preconceptions.

When looking at students' conceptual understanding it has been seen that the contextual framework of a question may have an effect on students' ability to apply a well known concept. A study by Savinainen and Viiri investigated the conceptual coherence of students using the FCI [144]. They noted that both the context and the diagrammatic or verbal representation of a question can affect students' overall performance. For a student to have a complete grasp of a topic they should be able to apply a physics concept to a variety of questions, irrespective of the surface features of the question or their representation. The FCI is a prime example of a test in which several questions focus on the same force concept but are represented by very different real life contexts.

Previous research into the use of free-body force diagrams by students to solve mechanics questions showed a positive correlation between students who drew diagrams and their success on physics problems [145]. Rosengrant *et al* noted that students who were exposed to a teaching environment in which multiple representations were employed, and an emphasis was placed on encouraging students to draw such diagrams, were more likely to choose to draw diagrams in their own problem solving strategy [145].

Presented in this chapter are results collected from a qualitative study undertaken to probe students' reasoning and understanding of force and motion. Qualitative analysis obtained from student interviews at three UK universities provided a greater insight into the misconceptions held by students. It was not within the scope of this research to undertake a full qualitative study for the whole first year cohort. Consequently, a representative sample of students from each university was engaged in the interview process, from which specific examples of students' misconceptions could be deduced. This study aimed to investigate how students approach conceptual physics problems and whether there is a gender difference to their reasoning. Details of the design and implementation of this qualitative study, alongside results from interviewees will be discussed. In addition to an evaluation of male and female students' understanding of each physics problem presented in the interview process, a discussion of trends in student problem solving behaviour as a whole will be presented, with reference to their confidence in their own comprehension. The use of diagrams as a problem solving aid is also explored.

5.1 Qualitative interviews

As outlined in Chapters 3 and 4, results from several consecutive years of first year undergraduate physics courses present a picture of a persistent and very noticeable gender gap in a conceptual test of Newtonian concepts of force and motion. Differences were found in both overall FCI scores and in the number of correct responses on individual questions. Results showed a consistency in the observed gender differences between the University of Edinburgh and other UK universities. Having identified particular test items which featured common misconceptions about key Newtonian concepts, it is important to learn more about the origins of these misconceptions. Specific differences between male and female students' misconceptions, or misconceptions held by the class as a whole, highlighted in this study may offer potential opportunities for the implementation of interventions in future courses.

Interviews with participants enabled a study of how students processed a physics problem, techniques they used during the problem solving process, as well as a measure of their ability to isolate the physical concepts being tested in individual questions. By conducting the interviews across the three institutions discussed in section 3.2, which have very similar course contents, more information could be gathered about students' conceptual understanding. This can also help to confirm the previous hypothesis that these differences are not institution dependent, but are consistent across different UK universities. In this section the methodology of the qualitative study undertaken to probe these concerns will be outlined, along with a discussion of the analysis of data collected using structured interviews.

5.1.1 Qualitative methodology

There are a number of qualitative analysis techniques and approaches that may be employed, the choice of which must be based on the objectives and aims of the conducted study. Qualitative data collection can take the form of focus groups, interviews, questionnaires or even an ethnographic approach. Such analysis is useful for gaining a more comprehensive understanding of a person's perspective of a topic or issue. Using such techniques more detail can be obtained than through analysis of purely quantifiable data. Parlett and Hamilton described qualitative analysis as taking

> *"account of the wider contexts in which educational programs function. Its primary concern is with description and interpretation rather than measurement and prediction."* [146]

A qualitative approach can provide greater focus on the techniques employed by individual students, which, when combined with the quantitative results for the whole cohort, may bring us closer to understanding the factors that make a difference in students' performance. For example, factors such as an individual's lack of domain knowledge about a specific concept can potentially be masked by a high quantitative test score which could be obtained solely through their ability to use formulae. The ability to interact with individuals through focus groups or interviews allows for the researcher to probe a student's understanding to a deeper level. A limiting factor of qualitative research is that it is not possible to carry out analysis on the whole cohort due to practical and time restraints, and, as in this study, results are restricted to a handful of students. It cannot be assumed that the views held by this sample are shared by the whole cohort.

5.1.2 Interview structure

The research study discussed in the following sections used interviews with individual students to look for patterns in the way in which students conceptualise different physics concepts of force and motion. Interviews are a widely used form of systematic inquiry. Interviews as a method of research aimed at gathering information through conversation, either with individuals or focus groups, can take many forms. Potter states that conducting interviews offers the benefit of providing

> *"an area for identifying and exploring participants' interpretative practices rather than an instrument for accessing a veridical account of something that happened elsewhere."* [147]

Unlike for quantitative surveys or assessments, where the information collected is limited by the instrument, interviews allow for an interaction between the interviewer and the participant and an opportunity for the questions posed to be personalised depending on the responses given by individual students.

In this study, interviews took place with individuals rather than focus groups. Although slightly more time consuming, individual discussion is less likely to be dominated by a single student in a group and would allow for more in-depth discussion about an individual's reasoning. It is important to recognise that, although generalisations about the whole student population cannot be made, results of these interviews provide an insight into students' approaches to conceptual questions as well as the misconceptions held by students. It was not in the scope of this project for qualitative analysis to be collected for an entire cohort, but rather a selection of students aimed to be representative of the population was chosen.

One of the important factors to be addressed in the methodology is the structure of the interviews that take place [148]: unstructured or structured interviews. Unstructured interviews enable the interviewer to ask open-ended questions of the participant, often resulting in a more personal and open account. These interviews do not follow a predefined structure or set of interview questions, allowing for a very relaxed atmosphere. Structured interviews do follow a predefined format in which the interview questions are decided prior to the interview taking place. This allows for a high degree of consistency to be maintained between consecutive interviews, and therefore for direct comparisons to be made between participants. Such interviews are often employed when investigating a specific quantitative or opinion response.

For the purpose of the study presented in this thesis, interviews followed a semi-structured format. A series of physics problems was chosen prior to the interviews and students were prompted to explain their choice of answer after the completion of each question, if they had not already done so during the think aloud process discussed in the next section. The semi-structured format allowed for deviations from set sequences of questions prohibited in a fully structured interview. The interviewer had to engage with active listening, altering their interview questions with respect to the response and physics working of each student. The interviewer used a series of evaluative and descriptive questions probing a student's ability to explain and evaluate their reasoning without unintentionally guiding the student through the problem by mentioning any physics concepts that may influence their preconceived ideas about the tested topics. Although interviewers had to personalise some of the questions as a result of each student's approach, it was important to ensure that the same questions, or range of questions, were posed to all participants in the study and that these were presented in a similar manner and style to ensure consistency [149]. An interview script, indicating the order of the physics problems posed to students as well as questions to be asked of the students, was created to maintain consistency across the three universities.

The first set of interviews, at the University of Edinburgh, was carried out in the presence of three instructors, one from each of the participating universities. Students were informed that only one researcher would be involved in conducting the interview, the others acting purely as observers to ensure the interview practice was repeated as identically as possible when conducted at the other participating institutions. In each case, to reduce any potential pressure felt by the student, the interviewer was not a lecturer at the university in which the interviews were taking place. Livescribe pens were used during the interview process as a method of recording [150]. This technology allows for students' written working and diagrams to be recorded in complete synchronisation with the audio recording. This method of recording was chosen as it allowed a complete record to be kept of students' working, including any diagrams they may have drawn during the interview. It was considered less intrusive than video recording and allowed for students' responses to flow more freely.

The interviewer had little input into the student's physics working, but their presence allowed for them to suggest that the student move on from one question to the next if they were lingering or stuck on a particular concept. The aim of this study was not to test whether the student answered the question correctly but to further understand how they chose their answer, and subsequently their reasoning for disregarding possible alternatives. Therefore, even if the student answered incorrectly, but had provided a full explanation of their reasoning, they were prompted to move on to the next question.

5.1.3 Think aloud interviews

Interviews as a form of verbal reports can take different forms; concurrent reports or retrospective reports [151]. The concurrent interviews carried out in this study used a think aloud technique. This technique is now well established for investigating strategies for problem solving [151]. Participants vocalise their thinking and problem solving approach whilst working through each individual question. This enables the researcher to gain additional information about the processes used by the students which were not provided by either written answers or numerical data. These indicate purely whether the student answered the question correctly or not.

It was important to ensure an interview environment in which students verbally expressed their thought processes as they approached the problem. Cottle discusses the importance of creating this environment for communication:

> *"Without allowing people to speak freely we will never know what their real intentions are, and what the true meaning of their words might be."* [152]

The process of getting participants to vocalise their thinking during the interview process will have an unpreventable effect on students' natural behaviour when solving the conceptual physics problem. Whether this has a positive or negative effect on their working is unclear. By being asked to vocalise their thought processes as they work through a problem, students are forced to consider their responses, and as a result their problem solving strategy may be affected. They may approach the problem in a more structured way than they would do naturally. The extent to which this affects their thought processes or cognitive load has been debated [151, 153, 154]. One consequence of asking students to vocalise their thinking is that their verbalisation may not be in synchronisation with their cognitive process, leading to an incomplete record of their problem solving strategy [153]. Some students also report that they found it difficult to verbalise all their thought processes during the interviews and that their thought processes are more complicated than could be related to the interviewer [153]. It is therefore important to be aware that it is unlikely that every element of their thought process will be verbalised.

The presence of the interviewer may also have an adverse affect on students' natural approach to a problem [155]. In some cases students may be influenced by what they believe the interviewer is expecting to hear and may be unwilling to offer a full explanation of their reasoning if they fear they are incorrect. Hammersley however discusses the need to recognise that the use of an artificial environment as a setting for a research study does not necessarily invalidate the data collected:

> *"While it is true that the participants' behaviour is often influenced by the experimental situation and by the personal characteristics of the researchers, this by no means renders the results of experimental research of no value. Much depends upon whether the reactivity involved affects the results in ways that are relevant to the research topic and in a manner that cannot be allowed for."* [156]

It was noted that students had an increasing familiarity with the interview procedure as time progressed, anticipating the need to explain their reasoning. Each interview focused on nine multiple choice conceptual physics questions from the FCI. The interviewer asked each student to verbalise their reasons for choosing each multiple choice answer, or equally, their reasons for eliminating multiple choice options they believed to be incorrect. It follows that some students began to anticipate that they would be asked for such explanations and consequently provided details without additional prompting. Students did show their ability to verbalise their thought processes. It could be argued that the interview process led to students becoming more metacognitively aware and acted as a prompt to encourage evaluation of their answers. In order to ensure that students were comfortable with the think aloud process, the first question was chosen as a warm up question and used to familiarise the participants with the use of the Livescribe pen and with the process of vocalising their problem solving strategies.

5.1.4 Interview participants

A total of 34 students participated across the three institutions. Six first year undergraduate students at the University of Edinburgh participated in the study in March 2012 along with 14 students from University of Manchester and 14 from University of Hull. Ethics approval and consent was gained from each university for participating students. Participation did not contribute to course assessment and no course credit was given.

It was important to recruit a selection of students across all performance levels. Students, who participated in both pre-test and post-test FCI diagnostic tests were randomly selected from each pre-test quartile. It was also imperative that the study included a sufficient number of male and female students to test the gender hypothesis. In total this study included 14 females and 20 males. Students were initially invited to participate through an open invitation via email, followed by targeted re-emailing to sections of the cohort not already represented. At Manchester and Hull students from under-represented quartiles were then approached directly to participate in the study during class time. It was not the case that at each university students from each of the four performance quartiles participated in the study. This was particularly the case at the University of Edinburgh in which no students in the bottom performance quartile volunteered to take part in the interviews. Although no course credit was given for participation, students were given vouchers for their participation. Each interview lasted between 25 and 30 minutes.

5.1.5 Coding of interview results

Qualitative data can be analysed using many different approaches: grounded theory, discourse analysis, semiotics, thematic analysis or content analysis [147, 157]. The grounded theory approach developed by Glaser and Strauss can be considered as a reverse engineered hypothesis. Categories or themes are developed through introducing codes as and when they emerge during the data collection and analysis process. From these codes categories are then formed and a theory developed from the relationship between the concepts. Grounded theory benefits from the ability to view the data from a new perspective, as mentioned by Charmaz [158], but is limited by the practicality of expecting researchers to prevent their awareness of current theories from influence their coding or categorisation of the data. Similar to grounded theory, thematic analysis involves the coding of data through categories which emerge from the data and are not imposed on it prior to analysis. However, it is more likely that researchers completing a thematic analysis develop categories or themes with reference to previous research or literature [159].

Although not used in the context of this thesis, discourse and semiotic analysis can be employed to investigate language and stylistic nuances in interviews. This form of analysis is particularly useful if information on a subject's use of language or interaction with others is particularly relevant to determining their opinion or attitude towards a specific topic [147].

After the completion of the interviews, four researchers from the participating universities listened to the recordings of a sample of the student interviews and initially coded them with a constructivist view, allowing different researchers to arrive at different, equally valid, theories through the analysis of the same data. From this open coding process a series of more detailed categories was created to look at several specific areas of interest:

- Gender
- Was the student's answer correct or incorrect?
- If answered incorrectly, what multiple choice answer was selected?
- Did the student change their answer after prompting from the interviewer?
- Did they draw a diagram?
- Explanations for choosing or eliminating multiple choice options
- Key physics concepts discussed
- Did the student make reference to a previously answered test question?
- Speed of response
- Confidence level

When coding the 'Speed of response' for each student on each of the test questions, their response was scored as having been answered 'Immediately' or 'After Deliberation'. Similarly, the confidence level of the students was marked by coders on a three point scale ('Very Confident', 'Somewhat Hesitant' and 'Very Unsure') in order to gauge an idea of how well the student felt they understood the problem and how secure they were in their choice of answer. When considering the validity of the data analysis, each interview was coded by two different interviewers using the categories shown above. Neither of these coders was the researcher conducting the interview being coded. In the event of a disagreement, discussion was undertaken until a consensus was reached about students' confidence or speed of answer.

5.2 Interview questions

Nine questions were selected from the FCI to be used in the interviews. Questions included both test items that indicated statistically significant gender differences at the 95% confidence level (Questions 8, 13, 14, 21 and 23) and questions in which both genders showed similar levels of performance (Questions 7, 22, 26 and 15). This allowed for conceptual understanding on a general level, as well as any potential gender issues, to be probed. In the following section trends and patterns in students' approaches and their answers to each of the questions will be discussed with specific reference to quotations taken from student interviews.

5.2.1 Question 7

7. A steel ball is attached to a string and is swung in a circular path in a horizontal plane as illustrated in the figure below.
At point *P*, the string suddenly breaks near the ball.
If these events are observed from directly above, which of the paths 1–5 below would the ball most closely follow after the string breaks?

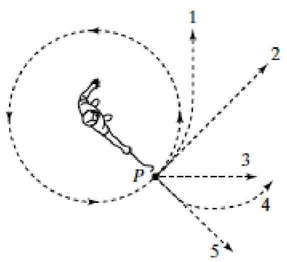


Figure 5.1: Question 7 from FCI. See Appendix A for complete version of test [46].

The first of the nine questions presented to the students was chosen as a warm-up question. This question, shown in Figure 5.1, had a very high percentage of correct responses by both genders and required no mathematical calculations. Question 7 asks students to decide the resulting direction of the motion of a ball, swung in a circle in the horizontal plane on a string, after the string breaks. Newton's First Law states that, in the absence of a net force, an object will continue to move in a straight line. In the case of the ball on the string, the ball will continue to move in the direction of the instantaneous velocity at the point where the string breaks (option 2).

Considering all students in the 2011-12 cohorts at the three universities, it was seen that 86% of students at Edinburgh answered correctly in the pre-test, compared to 83% of Hull students and 94% of Manchester students. A chi-squared test found no statistically significant gender differences in pre-test results at any of the three universities (Edinburgh χ^2=0.289, p=0.591, Hull χ^2=0.350, p=0.554 and Manchester χ^2=1.160, p=0.282). In each case a higher proportion of male students answered correctly than female students. The high percentage of correct responses allowed for this question to be used as a warm-up question to introduce the students to the idea of verbalising their thought processes.

Table 5.1: Summary of interview results for FCI Question 7. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	20 100%	13 93%	0 0%	1 7%	4 20%	2 14%
Edinburgh	4 100%	2 100%	0 0%	0 0%	2 50%	1 50%
Hull	9 100%	4 80%	0 0%	1 20%	2 22%	0 0%
Manchester	7 100%	7 100%	0 0%	0 0%	0 0%	1 14%

Results of the interviews are shown in Table 5.1. Of the thirty-four students who took part in the interview process, all but one female participant from Hull answered correctly. This question focuses on an object undergoing circular motion, with a velocity vector acting tangentially to the circle. Students should recognise that when the string breaks the ball is no longer acted on by a force into the center of the circle. The ball will travel in a straight line along this tangential path. This concept was well understood and explained by participants. One Manchester student referred to the tension on the string stating:

"If he releases it there is no tension force ... T [tension] *is zero and therefore there is no circular motion and it just continues its motion in the direction of* [path 2]*."*

The female who answered incorrectly indicated that the ball would follow path 1 after the string breaks. When providing reasoning for her answer she stated that:

"When you've got the string it's going to be accelerating and the force is going to be at right angles ... it's going to be at right angles to the string so every time it moves it is going to adjust so it's always at right angles ... So when the piece of string breaks it's going to fly off towards the direction that force is pointing at that point ... it will continue swinging round but because the force then stops, it will eventually continue in a straight line."

She correctly identified the presence of the centripetal force acting on the ball during its circular motion, but incorrectly assumed that the force will continue to act momentarily after the ball breaks contact with the string. She did comment that after some time the ball will eventually continue in a straight line, but associated this with a dissipation of the centripetal force. This misconception was reiterated when she was asked to discuss why she eliminated other possible options. Of particular interest was the statement:

"Not going to be 2 because as the string breaks there is going to be some extra force from where it snaps so the force is going to change."

When prompted she was unable to identify the origin of the "*extra force*". This line of reasoning is consistent with the misconception that the act of setting an object in motion supplies the object with some impetus or internal force, the magnitude of which dissipates over time if the body is no longer in contact with the source. This is sometimes referred to as 'impetus theory' [103, 160, 161]. In this case her statement reflects the belief that when the string breaks the ball maintains a curvilinear impetus and will continue on in its original circular path momentarily before becoming progressively straighter as this force gradually decreases.

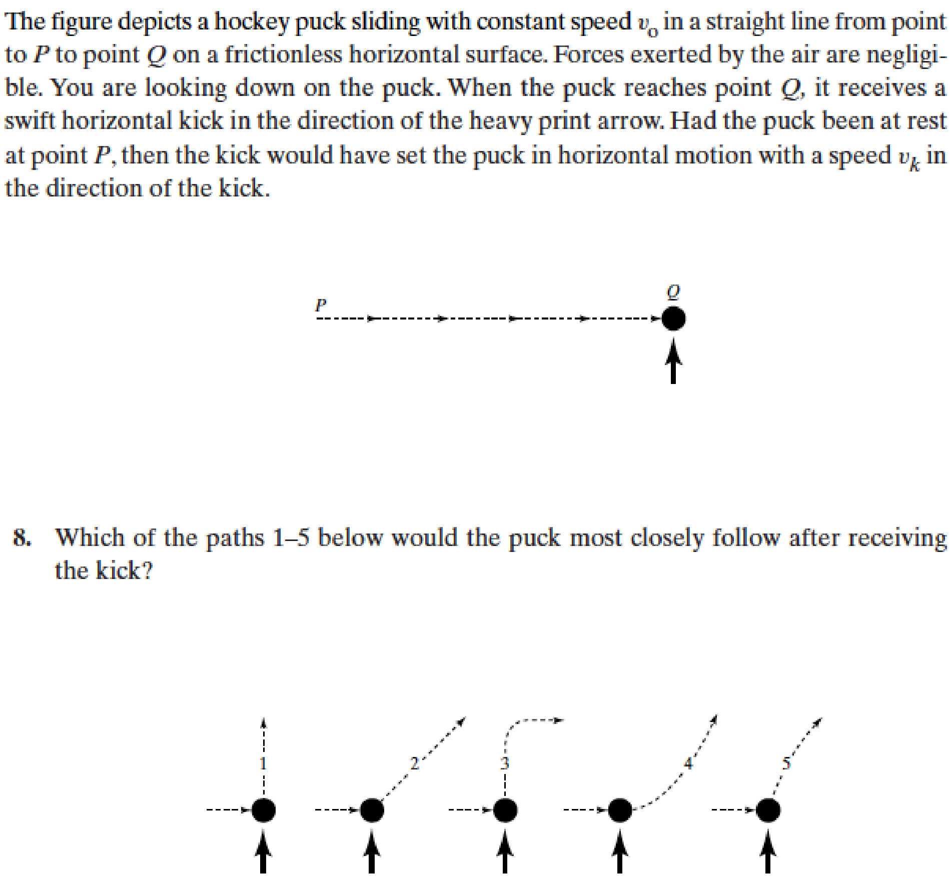
Students showed very high confidence in their understanding of this question, with 88% of students coded as being 'Very Confident' and answering 'Immediately' without any hesitation. This was not unexpected due to the high performance on this question by all three first year cohorts. Interestingly, the female student who answered incorrectly also answered the question immediately and showed a high level of confidence in her response. Gender differences in confidence levels will be discussed in greater detail in section 5.3.1. Although the question itself contained a detailed image, six students drew additional diagrams during their explanation. All of these diagrams involved arrows depicting the direction of the velocity vector associated with the ball. Each student drew the velocity vector as acting tangentially to the circle.

5.2.2 Question 8

Question 8, shown in Figure 5.2, contains two visual diagrams depicting the motion of a hockey puck sliding on a frictionless horizontal surface before and after receiving a kick. Students are asked to decide what motion the hockey puck will undergo after it has been kicked in the indicated direction. After being kicked in a direction orthogonal to its initial motion, the puck will follow a path in a straight line in the direction of option 2.

This test item was chosen as a question which demonstrated large gender difference across the whole class. In the 2011-12 pre-test results, the difference between the number of correct responses for males and females was statistically significant at both Edinburgh (χ^2=7.099, p=0.008) and Manchester (χ^2=8.275, p=0.004). In the pre-test, 65% of males and 40% of females at Edinburgh answered correctly and 77% of males and 57% of females answered correctly at Manchester. At Hull the gender difference was not significant (χ^2=0.588, p=0.443). As commented on in Chapter 3, it is not possible to derive clear conclusions from the Hull data as only six female students were included in the matched (pre- and post-test) data set.

Examining the results from the student interviews, shown in Table 5.2, a wide variation in responses to this question was seen. Of the fourteen female students participating, 57% correctly chose path 2. Of the six females who



The figure depicts a hockey puck sliding with constant speed v_0 in a straight line from point to P to point Q on a frictionless horizontal surface. Forces exerted by the air are negligible. You are looking down on the puck. When the puck reaches point Q, it receives a swift horizontal kick in the direction of the heavy print arrow. Had the puck been at rest at point P, then the kick would have set the puck in horizontal motion with a speed v_k in the direction of the kick.

8. Which of the paths 1–5 below would the puck most closely follow after receiving the kick?

Figure 5.2: Question 8 from FCI. See Appendix A for complete version of test [46].

answered incorrectly, three changed their response to the correct answer after being prompted by the interviewer to further explain their reasoning. The percentage of males who answered correctly was higher (85%), with one of the three incorrect students correctly changing their response during the interview.

This question tests whether students understand the effect of adding an external force to a system in a frictionless environment. The hockey puck initially has a constant horizontal velocity. After being kicked in an orthogonal direction students must consider the vector addition of the two velocities. The most common incorrect responses were paths 4 and 5. These two options depict curved paths which suggest that the puck is acted on by a force of changing magnitude. One of the male students from Manchester who chose option 4 stated that:

"... it should be 4 because ... as time goes on, because it's adding on a new

Table 5.2: Summary of interview results for FCI Question 8. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	17 85%	8 57%	3 15%	6 43%	6 30%	5 36%
Edinburgh	3 75%	1 50%	1 25%	1 50%	2 50%	1 50%
Hull	9 100%	3 60%	0 0%	2 40%	2 22%	1 20%
Manchester	5 71%	4 57%	2 29%	3 43%	2 29%	3 43%

acceleration, it's going to influence the direction more so, but it won't go completely vertical because there is still a horizontal force."

He was aware that on the frictionless surface the horizontal motion will be unaffected by the kick, but incorrectly stated that it is a force rather than a velocity. Interestingly, he commented on the fact that the direction of the puck will curve towards the vertical direction. This contradicted his statement about why he eliminated path 5 as a possible answer:

"because that would assume that the force upwards is decreasing, whereas it is constant because there is no frictional force on the upwards direction."

He commented that the curved path suggests a change in the magnitude of the force acting on the object. He goes on to state that the correct option, path 2, cannot occur because this shows a *"constant change in velocity"*.

A female student showed some confusion due to the diagrams themselves. Having initially indicated that she believed path 5 was the correct response, she stated that:

"Number 2 would depend on what the speeds were. If the speeds in both directions were the same then it would be number 2."

The interviewer pursued this line of thought and asked if path 2 was not pointing at a 45 degree angle, but instead was a straight line pointing more to the vertical, would the student be more comfortable choosing this as the correct answer. The female student agreed, elaborating that option 5 therefore indicated that:

> "*... it would be slowing down and if it was a frictionless surface it would be constant* [as in path 2]."

Another incorrect option chosen by two participants was path 1. This option in particular focuses on the misconception that motion requires an acting force. One interviewee stated that due to Newton's First Law

> *" if there's no other external force it has no reason to change in any direction ... I'd say 1 ... An object will just carry on in a straight line until it interacts with another force."*

Although correct that after the kick no external forces act on the hockey puck, she neglected to take the initial horizontal motion into account. Choosing path 1 suggests that she believes that the introduction of the force in the upward direction completely cancelled out any previous motion. There were some differences in the distractors chosen by male and female students. All the male students who were incorrect chose option 4, whereas females were more inclined to chose options 1 or 5.

Analysis of students' speed of response and confidence showed that students had much less conviction in their understanding than in the warm-up question. Compared to Question 7, in which most students answered immediately, over 62% of participants to Question 8 showed an element of deliberation when asked for the correct multiple choice answer. Similarly, the number of students coded as being 'Very Confident' was much lower than in Question 7, with 29% 'Somewhat Hesitant' and 9% 'Very Unsure'. Eleven students, five females and six males, drew diagrams during the interview process. All of these students answered correctly, excluding one who proceeded to change their response from a previously incorrect answer. These diagrams depicted arrows indicating the velocity vectors acting on the hockey puck. This demonstrated that they recognised that they needed to consider the addition of the two velocity vectors.

5.2.3 Question 13

Question 13, previously discussed in Chapter 4 (Figure 4.16), is a descriptive question describing a ball being thrown upwards out of a boy's hand. Students need to identify the forces acting on the ball as it is thrown upwards and subsequently falls back to the ground. The correct answer is option 4: *"an almost constant downward force of gravity only."*

Results from a whole class analysis of students across the three institutions in 2011-12 showed a statistically significant gender gap at Edinburgh (χ^2=5.625, p=0.018) and Manchester (χ^2=16.998, p<0.001). At Edinburgh 53% of males compared to 31% of females answered correctly. At Manchester pre-test scores were higher, with 79% of males compared to 52% of females answering correctly. Only one third of the female population at Hull chose option 4, compared to 61% of males (χ^2=0.685, p=0.408). Once again it is important to remember that only six female students were included in the matched data set for Hull.

Table 5.3: Summary of interview results for FCI Question 13. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	17 85%	8 57%	3 15%	6 43%	8 40%	6 43%
Edinburgh	4 100%	2 100%	0 0%	0 0%	3 75%	2 100%
Hull	7 78%	2 40%	2 22%	3 60%	3 33%	2 40%
Manchester	6 86%	4 57%	1 14%	3 43%	2 29%	2 29%

Table 5.3 summarises results from students' interviews for Question 13. Just over a quarter of the students interviewed answered incorrectly: six females and three males from Manchester and Hull. Almost all of these students chose option 3 which incorrectly states that, in addition to an almost constant force of gravity, there is an upward force acting on the object that steadily decreases until the maximum height is reached, and on the way down there is only an almost constant gravitational force. The popularity of this distractor, both in the overall cohort,

as seen in section 4.2, and in the interview analysis, suggests that a considerable number of students held the misconception that there is a lingering effect due to the initial upward force given by the thrower. This preconception is evident from the statement from a student from Hull:

"If the boy's throwing the ball up, it's going to have the upward force from throwing it up ... As it leaves his hand it will be slowly decreasing, then gravity will obviously be working on it ... It loses kinetic energy at the top ... It will go up and it will gain potential energy ... [it will gain] *kinetic energy as it falls back down ... Not number 4 because that's only talking about gravity. It's not talking about the force pulling it up."*

This female is confident that in order for the ball to move in the vertical direction it must be being acted upon by an external force during its motion and that the downward force of gravity cannot be the only force acting on the ball. This is similar to a response from another interviewee who chose option 2:

"For 4 there will be close enough to an almost downward force of gravity but it is not the only force that needs to be taken into account. When the ball is going up it's going to have a force applied to it once it's left the boy's hand but, due to the gravity, that upward force is going to decrease until it reaches its highest point where it will come to rest and then it will begin to accelerate towards the Earth due to gravity. So ... It's difficult because the only slight difference between the two [options 2 and 3] *is the difference in gravity so it will change slightly but at a very very accurate point so ... I'll go with 2."*

Once again the student stated that there must exist an additional force propelling the ball upwards. They commented on the difficulty of choosing between options 2 and 3. Unlike in option 3, option 2 suggests that the gravitational force does not affect the upward motion of the ball, but only acts during its downward motion. Some students had difficulty realising that the fundamental principle that gravity acts on an object even if it is travelling upwards. A male student from Manchester eliminated option 1 as a possible answer. When asked for his reasoning behind this decision, he stated that the upward force cannot be acting during the entirety of the ball's motion as the multiple choice answer suggests:

"It can't be 1 because at the top of the ball's trajectory ... ascent ... if there was a decreasing upward force it would carry on. It would only slow down. It wouldn't stop because at the top there's zero upward force. The resultant force overall is zero. So it should be number 3 because gravity is acting on the ball all the time it is going up. But the resultant force is upwards as the ball increases [in height], *but then as it slows down and reaches the top of its flight it's going to have a resultant force of zero and then at that point there's no upward force any more."*

Analysis of this question is key to understanding students' misconceptions about the relationship between force and motion. The above examples have highlighted that those students who answered incorrectly to Question 13 of the FCI believed that the act of throwing an object results in a continuing force acting on that object even after it has left the thrower's hand. In the absence of a net force, an object will remain at rest or continue at a constant speed. No additional force is required to maintain this motion. It addition to the existence of this force throughout the ball's motion, students stated that the ball slows down due to the magnitude of this force decreasing steadily during the upward motion, a misunderstanding also seen in the two previous questions. Although the majority of students understood that the downward force of gravity acts continuously on an object, two female students chose option 2, which suggests that it acts only on the downward trajectory, when the direction of motion is in the same direction as the gravitational force.

Similarly to Question 8, the speed with which students answered this question was very varied. Only 44% of participants were coded as having offered their answer 'Immediately', with the others deliberating and often systematically eliminating the other multiple choice options. When examining students' confidence in their responses, it was found that 68% of students were 'Very Confident' in their choice of answer. Only one male student, who answered incorrectly, was found to be 'Very Unsure'. This suggests that this question, and the relationship between force and motion, was found relatively difficult by students. None of the students who answered incorrectly drew a diagram. Overall 41% of students drew a diagram during the interview. The percentage of students who drew diagrams varied considerably by university. For example, it can be seen by Table 5.3 that almost all students at the University of Edinburgh used a diagram, compared to only 29% of students at Manchester. This may suggest an institutional difference in teaching or learning strategies. Students at Edinburgh are explicitly encouraged to include relevant diagrams in their written assignments and marking schemes reward students for doing so.

5.2.4 Question 14

Question 14 illustrates a scenario of a bowling ball being dropped from a plane travelling in a horizontal direction, as shown in Figure 4.14. The ball will follow a parabolic path with a constant downward acceleration due to gravity and a constant horizontal velocity due to its initial motion (option 4).

This question was included in the interview because it was representative of a question for which there were statistical gender gaps and, for females, previous results presented in Chapter 4 indicated that there existed one dominant multiple choice distractor. Results from the 2011-12 cohort indicated statistically significant differences between males and females at both Edinburgh (χ^2=9.058, p=0.003) and Manchester (χ^2=4.399, p=0.036), with males performing more highly than females in both populations. A chi squared distribution revealed that there was no difference between the performance of the two cohorts at Hull (χ^2=0.017, p=0.895).

Table 5.4: Summary of interview results for FCI Question 14. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort, who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	17 85%	11 79%	3 15%	3 21%	4 20%	4 29%
Edinburgh	3 75%	2 100%	1 25%	0 0%	2 50%	1 50%
Hull	9 100%	2 40%	0 0%	3 60%	1 11%	1 20%
Manchester	5 71%	7 100%	2 29%	0 0%	1 14%	2 29%

Interview results for Question 14 are shown in Table 5.4. Of the six students who answered incorrectly in the interviews, three were female and three were male. Two of the male students subsequently changed their incorrect response to the correct answer during discussion with the interviewer. One point of interest is the difference between popular distractors for males and females. Of the incorrect students, all three females chose the distractor option 2, whilst the three males chose distractor option 3. This was very different to results collected for the whole Edinburgh cohort between 2011-13 discussed in section 4.2. There it was seen that whilst almost all males answered correctly, with the remaining males split between answer options 1, 3 and 4, the predominant distractor for females was option 1. This discrepancy between the interview and whole-class results may be due to the selection of students participating in the qualitative study. As commented on earlier, it was not the case that students from each of the four quartiles at each of the universities volunteered to participate. In particular at Edinburgh no students from the lowest FCI pre-test performance quartile volunteered to take part in the study. In this respect, these students were a self selecting group. Nevertheless, analysis of students' reasoning for these answers can help to identify which misconceptions should be addressed.

One of the key issues highlighted by this study was students' confusion over the frame of reference in which this question was set. The diagram was misinterpreted by a large number of students. Several students asked the interviewer for clarification of whether the image of the plane referred to its initial position as the ball is dropped or whether it related to the position of the plane at the moment the ball hits the ground. When discussing why they eliminated potential distractors, one woman commented that option 2, in which the ball's path is vertically downwards, would suggest that the ball had no horizontal velocity when it was first dropped, which is the case if you were viewing this from the plane's frame of reference. This confusion of frames of references was seen in the explanation from a female student:

"Is the airliner in that position when the ball falls out? So it's not going to be 4 or 5 because that ball's not going to move faster than the plane ... I wasn't sure if it would drop straight down because if the plane is stationary it would. I know it's stationary compared to the plane, but in comparison to everything else the ball has velocity ... so is it just 2?"

Although confident that options 3, 4 and 5, which suggest some forward motion, could be eliminated, when explaining her reasoning for eliminating option

1 she showed some hesitation:

> *"I don't know why it would fall behind because before I was thinking that the plane had moved, so in comparison to the plane it would be further back, but at the instance it drops out I think it would be just straight down."*

As well as the relative position of the airplane with respect to the observer, students also discussed the velocity of the plane. One student commented that *"it* [the ball's position] *depends how fast it* [the plane] *is travelling."* Further analysis of the interviews indicated that the male student who incorrectly chose path 3, a straight line indicating constant horizontal and vertical velocities, failed to take into account the vertical acceleration. Initially he said that:

> *"From the plane's point of view it should see the ball falling downwards ... For the person standing on the ground you should see the ball going in the same direction as the plane* [path 3]*."*

After further discussion with the interviewer he quickly changed his answer to option 4, noting that gravity would cause it to accelerate downwards. Almost all students recognised that the only force that would be acting on the ball is the gravitational force. Interestingly this was not the case for the previous question where many students believed the ball being thrown up was still acted on by an upward force, even after leaving the boy's hand.

Those students who answered correctly demonstrated a high degree of confidence in their responses, with only three students coded as being 'Somewhat Hesitant'. Students were, however, more likely to deliberate over their answer than in previous questions. Less than half of the thirty-four participants answered 'Immediately', with many talking through each multiple choice option before choosing their final answer. Only eight students drew diagrams, four of whom answered correctly. The diagrams which were drawn indicated the directions of the velocity vectors acting on the ball and denoted the relative horizontal positions of the plane and the bowling ball.

5.2.5 Question 15

Students' understanding of Newton's Third Law is tested in Question 15 (Figure 4.19). Students are presented with a scenario in which a car pushes a truck while speeding up to a constant cruising speed. The car will act on the truck with a force equal to that which the truck exerts on the car (option 1), despite the fact that the whole system is accelerating.

Pre-test results indicated that this question was found difficult by the majority of the first year physics population. As discussed in Chapter 4, cohorts from the three universities showed no significant gender gaps, with the percentage of correct responses very low for both males and females. In 2011-12 only 22% of males and 22% of females at Edinburgh correctly noted that the car and truck would exert equal and opposite forces on each other whilst speeding up to cruising speed (χ^2=0.009, p=0.926). Pre-test scores were only slightly higher at Manchester; 34% of males and 30% of females answered correctly (χ^2=0.334, p=0.563). None of the female students at Hull correctly chose option 1 and only 17% of males picked this option (χ^2=1.167, p=0.280).

Table 5.5: Summary of interview results for FCI Question 15. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	16 80%	8 57%	4 20%	6 43%	11 55%	4 29%
Edinburgh	4 100%	2 100%	0 0%	0 0%	2 50%	2 100%
Hull	6 67%	3 60%	3 33%	2 40%	5 56%	0 0%
Manchester	6 86%	3 43%	1 14%	4 57%	4 57%	2 29%

Looking at the results from the interviews, shown in Table 5.5, ten of the thirty-four students interviewed answered incorrectly; four males and six females. All of these students chose answer option 3: *"the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car."* This was consistent with the results from the 2011-13 Edinburgh data set in which this was the dominant distractor (section 4.2).

Some students were quick to associate this question with Newton's Third Law. One female student correctly stated that when two objects are in equilibrium they push on each other with equal and opposite forces. Despite this she went on to state:

"When things were in equilibrium they were pushing on each other the same amount, so the same forces were acting upon each other ... I wouldn't say it's 1 because they're not in equilibrium. The car's accelerating ... car would need to be pushing on the truck more than the truck's pushing back on the car to move the truck ... I'm sure it's not right but I can't think why, but that's what I would instinctively say."

As well as demonstrating a lack of confidence in her reasoning, it was made clear that, because the car is accelerating, she believed that Newton's Third Law would no longer hold in this situation. She was more confident relying on her intuition than the physics principle. This misconception was very common amongst students. Participants tried to resolve the concept of Newton's Third Law with the acceleration of the system, thereby introducing the principle of Newton's Second Law. One male student who, after some deliberation, chose option 3 as the correct answer commented that:

"the force that the truck is pushing back on it can't be equal to it otherwise it wouldn't be speeding up ... Is it Newton's Second principle which is F=ma? So, because it's getting up to cruising speed means there's acceleration, which means there must be a net force in that direction."

Analysis of students' reasoning for this question has highlighted several key factors in students' misunderstanding. First, although students were able to state and recite Newton's Third Law, they showed a lack of understanding of when it is valid. The idea that Newton's Third Law only holds true if objects are travelling at constant speeds was common amongst students. If two objects in contact with one another are accelerating, as in Question 15, a large proportion of students correctly assumed that there is an unbalanced force, but incorrectly associated this unbalanced force with the reaction pair forces:

“It says that the car is still speeding up so, the total force in this direction [to the right] *should be greater than the total force in this direction* [to the left] *to make the net force on the system directed that way ... If they are equal then the net force in this car truck system is zero.”*

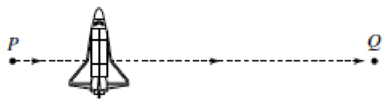
Student confidence was relatively low on this item in comparison to other interview questions. As in the examples above, several students were actively aware that they were missing a key element of understanding in order to resolve the fact that the car and truck were accelerating. Even those who answered correctly sometimes stated that they knew that Newton’s Third Law would hold for the car truck system but were unable to explain the resulting motion when asked for an explanation by the interviewer. In total nineteen participants were very confident in their answers, however, of these thirteen were incorrect. As well as demonstrating a lower level of confidence, students took longer to deliberate over their answers to Question 15 than in previous questions (56% answered ‘After Deliberation’). Eleven students used the Livescribe pen to draw a diagram during their discussion. These diagrams depicted the car and truck system with arrows denoting the direction of the forces acting on each body. Most students included the downward force of gravity and the upward reaction force. In many cases a larger arrow was drawn to indicate that the force acting forwards from the car was thought to be greater than the backward force from the truck. Students who answered incorrectly drew an incorrect free-body force diagram with the reaction forces acting on the wrong objects.

5.2.6 Question 21

Figure 5.3 shows a copy of Question 21 of the conceptual test. Question 21 is the first of four questions which discussed the motion of a spaceship after the addition and removal of an external force. Questions 21 and 22 were included in the interview process to lead students into Question 23, which showed a large gender gap in the first year cohort. In Question 21 a spaceship drifts sideways in space and is not acted on by any outside forces. After a time the spaceship’s engine turns on, producing a constant thrust in the vertical direction. Students are asked to consider the resulting motion of the spaceship. The correct answer (option 5) shows a curved path towards the vertical direction.

Looking first at results from the 2011-12 whole class populations, significant differences in the number of correct responses were seen between genders at Hull (χ^2=6.091, p=0.014) and Manchester (χ^2=8.143, p=0.004). No such gender difference existed in the Edinburgh cohort (χ^2=0.976, p=0.323). Once again males outperformed females at each institution.

A spaceship drifts sideways in outer space from point *P* to point *Q* as shown below. The spaceship is subject to no outside forces. Starting at position *Q*, the spaceship's engine is turned on and produces a constant thrust (force on the spaceship) at right angles to the line *PQ*. The constant thrust is maintained until the spaceship reaches a point *R* in space.



21. Which of the paths 1–5 below best represents the path of the spaceship between points *Q* and *R*?

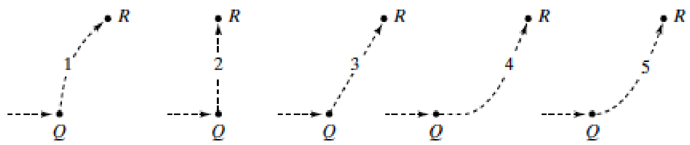


Figure 5.3: Question 21 from FCI. See Appendix A for complete version of test [46].

Results from the interview process are shown in Table 5.6. These indicate that fourteen of the participants answered incorrectly to the question. The proportion of incorrect responses was higher for the female population than for the male population. Despite this, nine of these students changed their answer to the correct one after further discussion. Those that did change their response had initially chosen path 4. This, along with supporting explanations from students, provided evidence for the hypothesis that students do not initially notice the subtle differences between the diagrams depicting paths 4 and 5. Path 4 suggests that the spaceship continues to travel for a short time in the horizontal direction before being affected by the engine's upwards thrust. Path 5 indicates that the

Table 5.6: Summary of interview results for FCI Question 21. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	13 65%	7 36%	7 35%	7 64%	4 20%	4 29%
Edinburgh	2 50%	0 0%	2 50%	2 100%	3 75%	1 50%
Hull	7 78%	2 40%	2 22%	3 60%	0 0%	0 0%
Manchester	4 57%	5 71%	3 43%	2 29%	1 14%	3 43%

effect of the engine being turned on is immediate. One student from Manchester stated:

"I would say it's path 4. This is similar to the hockey puck problem, but not quite the same because, although the spaceship is also moving with the velocity to the right, this time the thrust acts over a period of time versus the kick which is nearly instantaneous. So in this case it would form more of a sweeping curve."

The interviewer provoked further discussion by commenting that path 5 is very similar to path 4 and asked the male student to explain his reasoning for not choosing that option. The student then changed his response:

"Looking at it I may have to retract my statement and it is number 5 ... I guess I didn't look quite closely enough in that, for a period of time after the thrusters are on, it appears like the spaceship isn't moving upwards at all."

This proffers the question whether the recorded number of incorrect responses to this question is due to misinterpretation of the answer diagrams, rather than an underlying physics misconception. Are the results from this question testing students' conceptual understanding? Another student also commented on the similarities between the two multiple choice options, but argued that they could both be correct depending on the time delay for the thrusters to start up.

"Similar to the puck question, except instead of a quick application of force it is a force over a period of time ... At Q the engines are turned on, so it maintains its horizontal velocity. You might say that 4 could be just as right as 5, depending on how long it takes for the thrust of the engines to take effect."

Not all of the students who answered incorrectly chose path 4. One student who chose distractor 2 stated:

"If it's travelling from P to Q and there is constant thrust that's just going vertically up at a right angle it will move at a right angle up like in 2 because, like it says in the question, there are no outside forces and you don't need to take a resultant. You don't need to consider other forces apart from the one at a right angle."

The student believed that, once the engine is turned on, the original horizontal motion is no longer present but had been overcome by the upward force from the engine. This female student made a similar argument in Question 8 when referring to the motion of the hockey puck after it had been kicked. Another student chose path 3 as the path of the spaceship.

"It's going to be 3 because it is in space and there is no air resistance. The initial force that is acting on the plane to get it to go from P to Q is still going to be present, so that force is going to be constant because there is nothing resisting against it. There is no energy loss. So as soon as the force at Q is applied and it's continued to be placed on it, it's going to start deviating and the direction is going to change. It can't be 4 or 5 or 1 because it's gradual change which you would associate with the resistance."

They correctly assumed that the horizontal motion of the spaceship would continue, although they were unaware that there is no force acting on the spaceship in that direction, but that the motion is due to it drifting in space in the absence of friction. This shows that their misconception, that all motion requires an acting force, has been applied to each question they have attempted. They have assumed that a curved path must be associated with the presence of a resistive force.

An interesting result that emerged from these conversations is the frequency with which participants compared Question 21 with Question 8. Similarities were drawn between the horizontal motion of the spaceship and the motion of the hockey puck. Students correctly commented on the fact that, in the case of the spaceship, the force continues to act throughout the motion, whereas for the hockey puck the kick is instantaneous.

Students who chose an incorrect answer option during the interview, but did not change their response during the discussion, showed a high level of confidence. This was particularly the case for female students who were all coded as 'Very Confident'. The two male students who were incorrect were coded as 'Somewhat Hesitant'. Eight students, four females and four males, drew a diagram during the think aloud process and used this to explain their addition of the two velocity vectors.

5.2.7 Question 22

Question 22, shown in Figure 5.4, follows on directly from Question 21. It asks for a description of the spaceship's speed as it moves from point Q to point R. The correct answer (option 2) states that the speed will be continuously increasing because a constant force is being applied by the spaceship's engine.

22. As the spaceship moves from point *Q* to point *R* its speed is

___ 1. constant.
___ 2. continuously increasing.
___ 3. continuously decreasing.
___ 4. increasing for a while and constant thereafter.
___ 5. constant for a while and decreasing thereafter.

Figure 5.4: Question 22 from FCI. See Appendix A for complete version of test [46].

Chi-squared distribution tests showed no statistically significant differences in 2011-12 between numbers of correct responses for male and female students at either Edinburgh (χ^2=2.336, p=0.126) or Hull (χ^2=0.146, p=0.703). The Manchester population did show a statistical gender gap (χ^2=10.767, p=0.001) with 80% of males answering correctly, compared to 58% of females.

Results from the think aloud interviews (Table 5.7) indicated that this question was very well answered by students. Only four students answered incorrectly; three females and one male. Of these four students, two chose option

1 (constant speed) and two chose option 4 (speed increasing for a while and constant thereafter). The male student subsequently changed his answer to the correct answer (option 2).

Table 5.7: Summary of interview results for FCI Question 22. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	19 95%	11 79%	1 5%	3 21%	1 5%	1 7%
Edinburgh	4 100%	2 100%	0 0%	0 0%	0 0%	1 50%
Hull	8 89%	3 60%	1 11%	2 40%	1 11%	0 0%
Manchester	7 100%	6 86%	0 0%	1 14%	0 0%	0 0%

The majority of participants were able to identify that the concept of Newton's Second Law was being tested in this question, before moving on quickly to the next question. One student explained that:

"because the force is present the whole time from point Q to point R, which according to Newton's Second Law says that there will be an acceleration present, which means velocity will be increasing. It wouldn't be constant because there is a force present and it's not continuously decreasing because the force is acting in the direction of increasing velocity. It does increase, but it doesn't stop at all ... The thrust is present the whole time."

By choosing distractor option 1 students are demonstrating the misconception that despite the spaceship being acted on by a constant force, it continues at a constant speed. A female student who chose this option explained that, although there would be some initial acceleration, it would reach a constant speed almost immediately. One commented on the fact that the solution would differ for an object on Earth.

"because once the thrust ... but it's in space isn't it? ... I could definitely see why people would answer 4 ... because we live on Earth."

They referred to the fact that there exist resistive forces such as friction on Earth which are not present in this question. Option 4 suggests that after a while the object would stop accelerating and continue at a constant speed. These explanations highlight two existing misconceptions: that a constant force results in a constant velocity rather than a constant acceleration; presence of resistive forces in space.

Twenty-two participants (65%) were coded as showing high confidence in their answer, with a further eleven showing some hesitation. This is reflected in the fact that 62% of participants stated their chosen response 'Immediately'. Overall this question was answered relatively quickly by students. The use of diagrams was minimal. Only one female and one male student drew a diagram. This was not wholly unexpected since Question 22 referred to the same scenario depicted in the diagram of Question 21.

5.2.8 Question 23

Question 23, which was previously discussed in Chapter 4 (Figure 4.22), looks at the motion of the spaceship after its engine is turned off and the thrust immediately becomes zero. In answering this question, students need to consider the vector addition of two velocities. The correct answer is option 2.

Both Edinburgh and Manchester 2011-12 cohorts showed statistically significant gender gaps in the number of correct responses. At Edinburgh 78% of males answered correctly, compared to only 40% of females (χ^2=19.075, p<0.001). At Manchester 85% of males compared to 63% of females chose the correct answer (χ^2=12.580, p<0.001). No such statistical gender gap was present in the Hull population, with 75% of males and 50% of females answering correctly (χ^2=0.588, p=0.443).

Interestingly, thirty of the students interviewed correctly identified option 2 as the answer, much higher than results seen for the whole class cohorts at the three universities. Students who answered correctly made clear statements about the lack of force acting on the object after the thrust is removed:

"I'd say it's number 2 because now that the thrust is off there is no force on the space shuttle, which means that its direction can't change, so it has to continue in the direction of the x and y velocities."

Table 5.8: Summary of interview results for FCI Question 23. The table shows the absolute number of male and female students at each institution and the percentage of each university's gender cohort who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	20 100%	10 71%	0 0%	4 29%	2 10%	3 21%
Edinburgh	4 100%	2 100%	0 0%	0 0%	1 25%	1 50%
Hull	9 100%	3 60%	0 0%	2 40%	1 11%	0 0%
Manchester	7 100%	5 71%	0 0%	2 29%	0 0%	2 29%

All four students who were incorrect were female. Their responses were split between three distractors. The two students choosing options 1 and 4 highlighted the tendency of students to believe that after the force is removed the object will return to the direction of its initial motion prior to the force being added. Option 1, the most popular incorrect response for the Edinburgh 2011-13 female cohort, assumes that the motion changes instantaneously, corresponding to the instantaneously removal of the applied force. Conversely, option 4 suggests that the spaceship's change back to its original horizontal direction is more gradual:

"even though the thrust has dropped to zero it is still going to have momentum forwards so that will steadily decrease and then it will continue on sideways, the way it was going before."

The above explanation was provided by a student who went on to state that, after the engine is turned off, the spaceship will decelerate, therefore resulting in a curved motion. The student did not recognise that the question is set in the context of space, therefore in the absence of resistive forces. The idea of impetus or momentum was again voiced by another female student:

"when the engine turns off it's still going to be propelling forward and then it will go back [to its original direction] *and start drifting again."*

The overall high performance level by the interview participants is reflected in the large number of students coded as being 'Very Confident' in their answers.

Only five students showed some hesitation, three of whom were females who answered incorrectly. It was also noted that just under three quarters of the students immediately stated their chosen answer. Very few people drew or annotated diagrams to accompany their explanation. Only five students drew diagrams indicating the direction of the vertical and horizontal velocity components.

5.2.9 Question 26

A copy of Question 26 of the FCI is shown in Figure 5.5. In the preceding question a woman exerted a horizontal force on a box which moves along the floor at a constant speed. In this question students revisit this scenario and consider the effect of the woman doubling the force that she exerts on the box. The correct answer is option 5: the box will move *"with a continuously increasing speed."*

26. If the woman in the previous question doubles the constant horizontal force that she exerts on the box to push it on the same horizontal floor, the box then moves

___ 1. with a constant speed that is double the speed v_o in the previous question.

___ 2. with a constant speed that is greater than the speed v_o in the previous question, but not necessarily twice as great.

___ 3. for a while with a speed that is constant and greater than the speed v_o in the previous question, then with a speed that increases thereafter.

___ 4. for a while with an increasing speed, then with a constant speed thereafter.

___ 5. with a continuously increasing speed.

Figure 5.5: Question 26 from FCI. See Appendix A for complete version of test [46].

Although there existed no significant gender differences on this question in the pre-test at any of the three universities in 2011-12, the percentage of students answering correctly was low across both genders. For Edinburgh only 36% of males and 40% of females chose option 5 (χ^2=0.070, p=0.791). Pre-test scores were slightly higher at Manchester, 41% of males and 43% females (χ^2= 0.013, p=0.909). At Hull 8% of the male cohort and 17% of the female cohort answered correctly (χ^2=0.414, p=0.520).

Results from the interviews are summarised in Table 5.9. Although we can see relatively low percentages of correct responses for male and female students overall, there are noticeable differences between universities. All six students from Edinburgh answered correctly during the think aloud process. Results from Hull and Manchester were very different, with 22% of males and 20% of females correct at Hull and 43% of males and 57% of females correct at Manchester.

Table 5.9: Summary of interview results for FCI Question 26. The table shows the absolute number of male and female students at each institution, and the percentage of each universities gender cohort, who answered correctly, incorrectly or drew a diagram.

Institution	Correct		Incorrect		Diagram	
	Males	Females	Males	Females	Males	Females
All Students	9 45%	7 50%	11 55%	7 50%	11 55%	5 36%
Edinburgh	4 100%	2 100%	0 0%	0 0%	4 100%	2 100%
Hull	2 22%	1 20%	7 78%	4 80%	4 44%	0 0%
Manchester	3 43%	4 57%	4 57%	3 43%	3 43%	3 43%

The key factor being tested in this question is students' understanding of friction. There are two different coefficients of friction: static and kinetic. Static friction between two surfaces increases with increasing force applied up until the point of slip when motion occurs. When two surfaces are moving with respect to one another the friction is defined by the kinetic friction. The coefficient of kinetic friction is lower than that of static friction.

Students showed that they have a variety of misconceptions about friction. The most common distractors chosen by students were options 2 and 4, both of which suggest that eventually the box will travel with a constant speed. One student who chose option 2 commented:

"All the other answers assume that the speed will change but, if you assume that the force was changed immediately, the speed won't be changing. And obviously because it doesn't say that the horizontal floor was frictionless obviously the speed is constant due to the fact that there is friction between the box and the floor. If the change in force is not immediate there will be some intermediate stages ... and the ratio between the force and the friction force will change."

In this example the student has assumed that the presence of the frictional force balances out the force applied by the woman, resulting in a constant speed.

This idea that an 'equilibrium' stage is reached, in which friction eventually balances the forward force, was very prevalent in the interview results. This suggests that students did not have a complete understanding of the factors affecting the magnitude of the frictional force. This was once again seen in an explanation by a student from the University of Hull:

"You increase the force and it's not an instantaneous acceleration ... well it accelerates constantly because the force increases and as the acceleration increases so does the friction until it comes to an equilibrium point where the friction counteracts the acceleration and you move at a constant speed."

Analysis of the interviews showed that students also had an inconsistent view of the relationship between applied force and its effects on velocity and acceleration. One female student who chose option 2 (box travels with a constant speed not necessarily twice as great as v_o) stated that in order for the box to accelerate a constantly increasing force was needed.

"There may be resistances on the box, which means more force is required to push the box at the same speed or a greater speed. If it was a continuously increasing speed the woman would have to be exerting a constantly increasing force."

Here she has confused acceleration with velocity. By applying Newton's Second Law it can be seen that, as the net force acting on an object is increased, the acceleration will also increase. Therefore, by applying a constant force the speed of the object will be continuously increasing. Once again, this misconception was noted in an interview with a male student who stated:

"It can't be 5 because in order to make something travel with increasing speed you would have to give it a force that's also increasing. And then 4 ... it might be 4 actually because it's not going to be instantaneous when she pushes it. Does it just double straight away? Whether or not, when she pushes there's going to be like a certain amount of time the box increases and reaches like an equilibrium period."

Several students approached this question by trying to relate it to their own experiences. Instead of explaining their reasoning through discussion of physics concepts of force or friction, they compared the context of the question to real life situations. For example, one student said that they also took this approach for similar questions:

"Because like in all the dynamics stuff that we've done I've tried to like imagine what would happen if I tried to do it. Like if you push something there is more resistance on it for a while and then it kind of carries on normally."

Here, although he has not referred to specific physics terminology, he noted that when you begin to push a still box you need to overcome the static friction. Once the object starts to move the force exerted on the object needs to be greater than the frictional force, which is proportional to the coefficient of kinetic friction (where the coefficient of kinetic friction is smaller than the coefficient of static friction).

Only three students were coded as having answered 'Immediately' to the question. The vast majority of the students took a lot of time to deliberate, working through each multiple choice option before selecting their answer. It was also seen that, across the three universities, students had varying degrees of confidence. Fourteen students were coded as being 'Very Confident' with a further sixteen being 'Somewhat Hesitant'. Almost half of students drew a force diagram. The number of students drawing a diagram varied both by gender and university. All students at Edinburgh used a free body force diagram to explain their reasoning, and subsequently answered correctly. At Hull four of the male students drew a diagram, but none of the female students did. Interestingly, only one female student from Hull answered correctly. At Manchester 43% of both the male and female cohorts used diagrams. Only one of these six students did not answer correctly. This suggests that, for this question, drawing a diagram aided most students in their problem solving, in particular females.

5.3 Analysis of qualitative interviews

In addition to analysing the interviews on a question by question basis, results were explored for any underlying gender trends in students' confidence, speed of answer, use of diagrams and association between questions. In this section each of these subjects will be discussed with reference to comments made by participants.

5.3.1 Student confidence and speed of answer

For each question in their interview a student's response was coded for the confidence level they had in their answer. Table 5.10 shows the number of correct and incorrect responses as a function of their confidence level for male and female students. Only students who conclusively answered correctly or incorrectly were included in this analysis. Students who changed their initial choice of answer during discussion with the interviewer were excluded. As discussed earlier, the correct and incorrect answers were coded as either 'Very Confident', 'Somewhat Hesitant' or 'Very Unsure'.

Table 5.10: Percentage of correct and incorrect male and female responses to interview questions as a function of their confidence level. N(males)=20 and N(females)=14. The total number of responses for males was 168 and for females 116.

Answer	Gender	Confidence	Number of Responses	% of total cohort responses
Correct	Males	Very Confident	124	73.8
	Females	Very Confident	58	50.0
	Males	Hesitant	22	13.1
	Females	Hesitant	23	19.8
	Males	Very Unsure	1	0.6
	Females	Very Unsure	0	0.0
Incorrect	**Males**	**Very Confident**	**1**	**0.6**
	Females	**Very Confident**	**16**	**13.8**
	Males	Hesitant	14	8.3
	Females	Hesitant	13	11.2
	Males	Very Unsure	6	3.6
	Females	Very Unsure	6	5.2

Female students appeared more likely than males to answer incorrectly: 30.2% of female responses were incorrect, compared to 12.5% of male responses. This is reflective of the overall gender gap in favour of male students seen in the first year courses at the three participating universities. Despite answering incorrectly, females showed a higher degree of confidence in their answers. Only one incorrect response (0.6%) from a male student was coded as 'Very Confident', compared to sixteen responses (13.8%) by females. Similarly, male participants were more likely to show high confidence levels in their correct responses compared to

Table 5.11: Percentage of correct and incorrect male and female responses to interview questions as a function of the speed of their answer. N(males)=20 and N(females)=14. The total number of responses for males was 168 and for females 116.

Answer	Gender	Speed of Answer	Number of Responses	% of total cohort comments
Correct	Males	Immediately	86	51.1
	Females	Immediately	44	37.9
	Males	After Deliberation	61	36.3
	Females	After Deliberation	37	31.9
Incorrect	**Males**	**Immediately**	**0**	**0**
	Females	**Immediately**	**14**	**12.1**
	Males	After Deliberation	21	12.5
	Females	After Deliberation	21	18.1

females: 73.8% of responses from males who answered correctly were coded as 'Very Confident', compared to 50% of female responses. This may suggest that males are not only more likely to answer correctly, but also more aware of when they are incorrect, consequently being less likely to show a high level of confidence when prompted for their answer. Another possible explanation may be that females are more likely to choose an answer based on their first instinct, or one which coincides with their prior conceptions of a topic. If unsure of their answer male students may take more time to deliberate and consider every possible option before voicing their final answer.

Gender differences were also explored for those students who decided to change their answer from a previously incorrect answer during the think aloud process. This occurred twelve times for males and ten times for females over the course of the nine questions. Once again male students showed higher confidence in changing their answer: 58% of changed responses for males were coded as having a very high confidence level, compared to only 20% for females.

Table 5.11 shows results of the coding of the speed of students' answers. A slightly higher percentage of males were recorded as providing the correct answer 'Immediately' (51.1%) than females (37.9%). A key point of interest is the gender differences for incorrect responses. All incorrect responses provided by male students were given after an element of deliberation or following a conversation with the interviewer. In direct contrast to this, 40% of females' incorrect responses were stated 'Immediately'. It should be noted that these incidences did not all occur during one specific test question, but were split across all nine interview questions and involved six of the participating female students. This suggests that, although female students were more likely to get the question incorrect, they were also more likely to be confident in their reasoning and therefore offer an answer with less hesitation. Male students were more likely to deliberate over their responses, and this may in consequence lead to the higher percentage of correct answers witnessed for males. A potential factor that needs to be considered when drawing conclusions from both students' confidence and the speed with which they answered, is that for all participants this was the third time they had seen these FCI questions, having answered the pre- and post-test during the previous semester. A consequence of this may be that some students recognised the questions and remembered the answers they had previously submitted, rather than working through the problem during the think aloud process.

5.3.2 Use of diagrams

The use and annotation of diagrams is an essential tool in problem solving in physics. As commented on at the beginning of this chapter, previous research indicated that using free-body diagrams to solve mechanics problems showed a positive correlation with students' success on physics problems [145]. When students first begin their undergraduate physics degree emphasis is placed on the correct construction of such diagrams. This is particularly the case at the University of Edinburgh where students are actively encouraged to draw free body force diagrams when answering physics problems. As mention earlier, a Livescribe pen was made available to students at the start of the interview and they were told to use the pen and notebook provided for any working or diagrams they wished to draw. It was hypothesised that students who used diagrams as part of their process to conceptually understand a physics problem would score more highly than those who failed to use any type of diagram. The majority, but by no means all of the students, drew some form of diagram or equation at some point during their interview. The percentage of students drawing diagrams ranged from 6% for Question 22 to 47% for Question 26. One female student stated that:

> *"I always like to draw a picture or have a picture in front of me like this so I can see the arrows and I can see the direction. I can see which way the velocity's going, which way the force goes."*

Looking at the correlation between answering the FCI question correctly and drawing a diagram, it was seen that those who did draw a diagram were more likely to subsequently answer correctly, however no gender differences were seen. Thirty instances in which a female student drew a diagram resulted in a correct response, compared to one instance which resulted in the incorrect multiple choice answer being chosen. A similar pattern was seen for males. Forty instances in which a male student drew a diagram resulted in a correct response, and seven times it resulted in an incorrect response.

The use of diagrams was further investigated by considering those drawn after prompting by the interviewer. It was queried whether being prompted by the interviewer to draw or annotate a diagram would result in a student changing a previously incorrect response to the correct multiple choice option. In fact this occurred only six times; three times for a female student and three times for a male student. As commented on in section 5.1.3, the presence of the interviewer in conjunction with having to vocalise their thought processes may affect students' natural problem solving processes [151, 153, 154]. It is therefore difficult to conclude definitively whether diagrams drawn during the interview process were drawn as part of the student's natural problem solving strategy or whether they were drawn purely for the benefit of explaining their reasoning to the interviewer as part of the think aloud process.

The use of diagrams to answer interview questions appeared to be very question dependent. For example, only six students chose to draw a diagram for Question 7, a question answered correctly by almost all students. Interview transcripts suggest that those who drew a diagram did so purely for the benefit of explaining their reasoning for choosing their selected answer, or for eliminating other options, to the interviewer. These students had already chosen and vocalised the correct response prior to the diagrams being drawn. The two questions for which drawing diagrams was most popular were Question 15 and Question 26. For Question 15, 44% of participants drew a free body force diagram and for Question 26, 47% of students drew a diagram. These two questions are representative of FCI test items in which students need to consider the direction and magnitudes of forces acting on an object. They were also questions which showed a relatively high number of students deliberating over each multiple choice option before choosing their answer.

As mentioned previously, the diagrams accompanying a test item were a source of confusion in some cases. This was particularly the case for Question 14 which depicted a bowling ball falling out of a plane. One student commented that the answer could not be options 3, 4 or 5 because the ball could not travel faster than the plane. The diagram is drawn such that the plane is in its position at the point at which the ball is dropped. The potential paths drawn refer to the motion of the ball as it hits the ground. Some students misinterpreted this as the position of the ball relative to the plane at the end of the motion, consequently eliminating the correct answer. The diagrams in the multiple choice answers for Question 21 also caused some confusion. Students did not easily recognise the difference between path 4 and 5 of the spaceship. Many students chose answer option 4 before noting that this indicated that the spaceship continued in the horizontal direction before being acted on by the force from the spaceship's engine. Both these examples highlight the possibility that students' low performance on a particular test item could be a consequence of their misinterpretation of the question or accompanying diagram, rather than a lack of conceptual knowledge.

5.3.3 Association between questions

There was evidence of participants referring to similarities between interview questions. Students sometimes used these similarities with a previously attempted question to determine which physics concepts were relevant to the question at hand. For example, several students drew comparisons between Questions 8 and 21. Question 8 refers to a hockey puck sliding at a constant speed in the horizontal direction before being kicked in an orthogonal direction. Question 21 also describes an object drifting at a constant speed. In the later case the object is in space and the constant force applied is maintained during its motion. One student commented that these were the *"same question"* except that in space there is no gravitational field. The similarities between the two questions did sometimes cause the participants to interpret Question 21 incorrectly. By focusing on the surface features of the question, the fact that the force from the spaceship's engine acts continually was sometimes neglected. Although this occurred with very few students during the qualitative study, this may help to explain those students in the whole class cohort who chose the spaceship's path to be in a straight line similar to that of the hockey puck.

Chi *et al* discussed the categorisation of physics problems by surface features or underlying concepts by experts and novices [162]. Novices are described as being more inclined to categorise problems based on surface features, such as the context in which they are set. For experts, categorisation is primarily based on the physics concepts required to solve each problem. By associating one question with another, students can often misinterpret the required physics law applicable to a problem, and this in turn can result in an incorrect answer. The effects of the context of a problem on students' performances have been examined in previous studies [72, 144]. Huffman and Heller suggested that students may perform better if they are more familiar, or have had more real life experience, with the context in which the question is set [163].

5.4 Chapter discussion and summary

Results from qualitative analysis of students' responses to a selection of FCI questions have highlighted several misconceptions associated with Newtonian mechanics. In this chapter these misconceptions have been examined through interviews focusing on nine test questions. Although pre- and post-test FCI results presented in the previous chapters have shown the existence of such misconceptions, by undertaking interviews with individual students a clearer idea of both their preconceptions and interpretation of the questions was gained. The role of the FCI is to act as a measure of students' level of understanding of Newtonian mechanics and to draw attention popular misconceptions.

In a study by Rebello and Zollman, looking at the effect of distractors on the performance of students on an algebra-based introductory physics course, they found differences between students' responses to open-ended questions and FCI distractors [135]. Despite there being no differences in the percentage of correct responses on the two versions of the FCI questions, they noted that a significant proportion of the open-ended responses did not correspond to any of the multiple-choice answer distractors on the original test. This may suggest that, although the FCI allows for some misconceptions to be brought to the attention of the instructor through the popularity of incorrect answer options, it may mask other potential conceptual problems. As well as the underlying misconceptions, it is extremely difficult to discern where or in what context the preconception was created. There are a wide range of sources, such as prior teaching or personal experience, from which these could be manifested, but it is very difficult for them to be traced back to the origin of conception [131, 132].

Students have shown that there are areas in which they have difficulty associating the prescribed physics problem with 'real world' situations. For example students may accept, and correctly state, certain physics principles, such as Newton's Second Law, but have difficulty applying it to everyday experiences. When discussing the problem of the woman pushing a box on a horizontal floor, one student commented that:

"If you push a box in real life on a floor it doesn't continue to get faster and faster. So friction must increase as she increases the force on the box."

They identified what they believe to be the correct answer, but remain unconvinced that this situation would occur outside the context of the physics problem. This may stem both from a lack of domain knowledge or from a culture of physics problems being set in 'ideal' conditions which often ignore resistive forces.

One hypothesis for explaining the gender gap in performance is that females are more likely than males to place an emphasis on 'real world' connections in their understanding of concepts of force and motion, which form the basis for common misconceptions. This hypothesis may explain the differences in popular FCI distractors between males and females observed in Chapter 4. The relatively small number of female students taking part in the qualitative study, and the difference between responses from participating students and the whole first year cohort, do not allow for this hypothesis to be conclusively confirmed or rejected. There was however evidence to suggest that some students were aware of the incoherence of their conceptual understanding. Having commented that if they themselves carried out the motion in the question they would witness a different outcome, they nevertheless chose the correct answer which they knew to be true, but could not explain why.

The students participating in the interviews had previously studied all the concepts involved in each of the problems set and had in fact both seen and answered these questions approximately twelve weeks earlier. This suggests that they do possess the required knowledge. The students' difficulty in answering the questions at the time of the interview may be a result of an inability to access stored knowledge, or indeed retain it over that period of time. In order to fully solve these problems students may require the transfer of knowledge from the long term memory to the short term working memory. This poses the question that if given a longer time interval, would the student be able to solve these problems? To probe whether this is the case students could be given a longer time frame to consider the questions.

When attempting questions some students showed a desire to compare different questions to one another. One potential problem with the use of association when conceptualising physics problems is the potential for students to concentrate on the surface features of the problem. Laurillard stated that:

"from the student's point of view, the problem situation is not just the content of the problem as given but includes also the context in which it is given." [164]

This categorisation of questions by surface features, rather than the underlying physics concept on which the problem is based, has been noted as a key difference between an expert or novice problem solver [162]. By comparing physics problems in this way students run the risk of misinterpreting the problem and employing the incorrect physics to the situation. When discussing the approach they take to solving physics problems, one student commented that:

"The first things I look out for are constant speed or, if it was a different question, constant acceleration are the first two things that I look out for to see what equations I can use and what equations I can't use."

This suggests that, for some students, identifying potential equations and numerical values is the first thing they do when approaching an unseen problem. Once again, this can potentially result in students neglecting important information in the question. By being more aware of this strategic approach instructors could emphasise the benefits of identifying physics concepts before attempting a question.

Students' confidence in their answers was also examined during the interview process. As well as showing the strength of their conviction in their choice of answer, their confidence can also affect their use of physics terminology. Ideally a high confidence level should be correlated with a correct answer and a lower confidence level with an incorrect response. A study of experts and novices in the physics department at the University of California at Berkeley investigated the link between performance and confidence of these two populations on questions regarding acceleration [165]. Experts comprised male faculty of the physics department. Novices were a selection of male and female undergraduate students. It was seen that 77% of experts showed high confidence when giving a correct answer and only 5% showed high confidence for an incorrect answer. Interestingly, for the undergraduates the percentage of students who were both correct and confident (23%) was almost the same as the percentage who were both incorrect and confident (24%), suggesting that they may have been less aware of the gaps in their understanding of acceleration concepts. The relationship between performance and students' attitudes has also been explored in other science disciplines. Felder *et al* noted that women entered engineering courses with a higher level of anxiety and lower confidence than males [166]. As the course progressed their initial high expectations of their ability to perform well in assessments decreased. Interestingly, results from interviews at the Universities of Edinburgh, Hull and Manchester showed that female students were much more likely than male students to show a high confidence level in an incorrect answer. They were also more likely to state their incorrect answer 'Immediately' compared to incorrect male students, all of whom answered 'After Deliberation'.

There are limitations to the conclusions that can be drawn from this qualitative study. While at the Universities of Hull and Manchester students were approached directly to take part in this study during class time, participants from Edinburgh were all volunteers. Therefore the sample from Edinburgh represented in this research, although comprised both male and female students, were a self-selecting group of students, having volunteered to take part. There was difficulty in finding students to participate from the lower quartiles, which in Chapter 4 showed the largest degree of confusion in the answer response profiles. This was particularly the case at Edinburgh in which no students in the lowest quartile participated.

Whilst the quantitative results presented in the multiple choice response profiles of FCI questions in Chapter 4 indicated which distractors were most popular amongst students, and in some cases which distractors were more popular for males or females, undertaking qualitative interviews enabled for students' reasoning to be better understood. This method isolated common misconceptions surrounding the topic of Newtonian mechanics and indicated certain questions in which students may answer incorrectly due to the wording of the question or confusion with the accompanying diagram. It also enabled a measure of students' confidence in their understanding to be made, something that cannot be achieved through purely quantitative data collection. The results presented in this chapter highlight areas in which future interventions could be implemented to reduce the prevalence of certain misconceptions of Newtonian mechanics.

Chapter 6

Gender Differences in Performance on Course Assessments

Students can be assessed in a variety of different ways over the course of their degree. Studies have suggested that genders may perform differently depending on the form of the assessment administered, for example whether they are assessed through coursework or examinations [3, 65, 70]. There exists a view that female students may exhibit better performance than male students on continual assessments such as weekly coursework, for which there is often the possibility to discuss the work with peers in a group environment prior to submitting the assessment [65]. Previous studies have shown gender differences in performance in both coursework and examinations at undergraduate level [3, 56, 62]. For each of the seven semesters of an introductory mechanics course Kost *et al* found that females outperformed males on homework and participation scores, while males outperformed females by an equivalent amount on examinations [3]. Similarly, Docktor and Heller found that whilst there was no overall gender difference in course grade, males did have marginally higher scores when examination results were considered separately from lab reports and participation scores [62]. This gender difference in coursework performance has also been examined at secondary school level. Elwood showed that females had higher mean coursework marks in GCSE English, Mathematics and Science [66]. A study undertaken at Sussex University with over 600 undergraduate students sampled from a variety of disciplines showed that females performed better than males on both coursework and examination components of courses [68]. In addition to quantitative scores, student surveys indicated that girls expressed higher anxiety about their overall performance and were more likely to report themselves as being adequately prepared for assessments. Such differences merit further investigation and could affect what teaching methodologies are implemented in the future.

In this chapter results from analysis of coursework and examination data from undergraduate physics courses at the University of Edinburgh are discussed with respect to students' gender. In the preceding chapters evidence has shown that there exists statistically significant differences between male and female performance on the FCI diagnostic test. This, along with subsequent qualitative analysis, indicated the presence of misconceptions amongst students in questions testing Newtonian concepts of force and motion. This topic makes up only part of the syllabus of the introductory physics course. Therefore, by examining the end-of-course assessment results, it can be determined whether this gender gap is specific to the Force Concept Inventory and representative of students' understanding of Newtonian mechanics or whether it presents itself in other areas of the course syllabus. Furthermore, results for additional first and second year courses may establish whether these gender discrepancies persist after the first year of study. In order to fully understand the effect of assessment types, data was collected from core courses in each year of the undergraduate degree.

Physics has a particularly low proportion of females studying in the undergraduate programme in comparison to many other science and engineering courses, although they also remain greatly under-represented in computer science and engineering courses [167]. In particular, biology and chemistry have much higher levels of female participation. In this chapter results from chemistry and biology courses at the University of Edinburgh are presented and used for comparison with those from physics to investigate whether the gender demographic of the classroom setting has an overall effect on differences in gender performance. Literature has alluded to the idea that minority groups, such as gender, may demonstrate lower performance and lower task activity when they are integrated into an environment in which they are outnumbered [168, 169].

Potential gender differences in assessment types also provided motivation for a wider study of the extent to which instructional methodologies may affect gender performance gaps. The use of peer discussion and personal electronic response systems in lectures have previously been shown to have a positive effect on students' learning [50, 170]. Student participation and performance on Peer Instruction questions in first year lectures will be discussed and gender differences explored for individual questions.

In this chapter coursework, laboratory and examination results from first year physics courses at the University of Edinburgh will be presented alongside a comparison of the performance of male and female students between 2006-13. The magnitude of the observed gender gaps in each of these academic years will be discussed for each form of course assessment. How the gender performance gap changes over the course of the degree programme is explored, firstly by looking at student performance in core second year courses and then through a fully longitudinal study of two year groups of physics students. As mentioned above, undergraduate physics courses have very different gender profiles to other STEM subjects. In this chapter, a comparison is made between gender differences on the physics programme and those witnessed in first year chemistry and biology courses. Finally, the use of Peer Instruction in introductory physics lectures is discussed and student performance on in-lecture questions presented.

6.1 Fully longitudinal and pseudo-longitudinal studies

When comparing the performance of students in different years of the undergraduate programme, results can take the form of either a pseudo-longitudinal or a fully longitudinal study. The term pseudo-longitudinal describes the method of comparing the performance of one year group with another year group to provide a 'snap-shot' of the gender situation across the different years of the degree course. The pseudo-longitudinal method makes the assumption that students in each year group are the same as in all consecutive year groups in the study. Although the structure and content of core courses remained relatively constant over the past few years, there existed some instances in which new material and teaching methodologies, particularly in first year, have been implemented. This may have had a subsequent effect on students' performance in later years of their degree. This was discussed by Singer and Willett when planning a study of teachers' careers [171]. They commented that cross-sectional data reveal nothing about 'change' and do not take into account changes in work experiences that may be a result of different background characteristics or administrative changes dependent on the year they entered teaching.

Another approach to examining gender differences in student coursework and examination performance is to consider the progression of the same individuals over the entirety of their degree programme. This methodology is been referred to as a fully-longitudinal study. By undertaking a fully-longitudinal analysis, the need for the assumption that each year's cohort is similar is eliminated. One disadvantage of completing a longitudinal study is the length of time required by the study. A fully-longitudinal study of assessment performance was conducted for two year groups (students who entered first year in 2006-07 and those who entered in 2007-08).

6.2 Gender performance in first year physics

As discussed previously, there exist statistical differences between male and female scores on a test of conceptual understanding of Newtonian mechanics, both at the beginning and end of the introductory first year physics course. The content of the FCI is covered in the first five weeks of the eleven week semester and is therefore not the sole focus of the end-of-course examination or coursework assessments. Because of this, it is interesting to explore whether the observed gender gap in Newtonian mechanics is replicated in final scores for each of these assessment types. In addition to the introductory first year course, assessment results were collected for students between 2006-13 in the second semester first year course.

6.2.1 First year physics courses

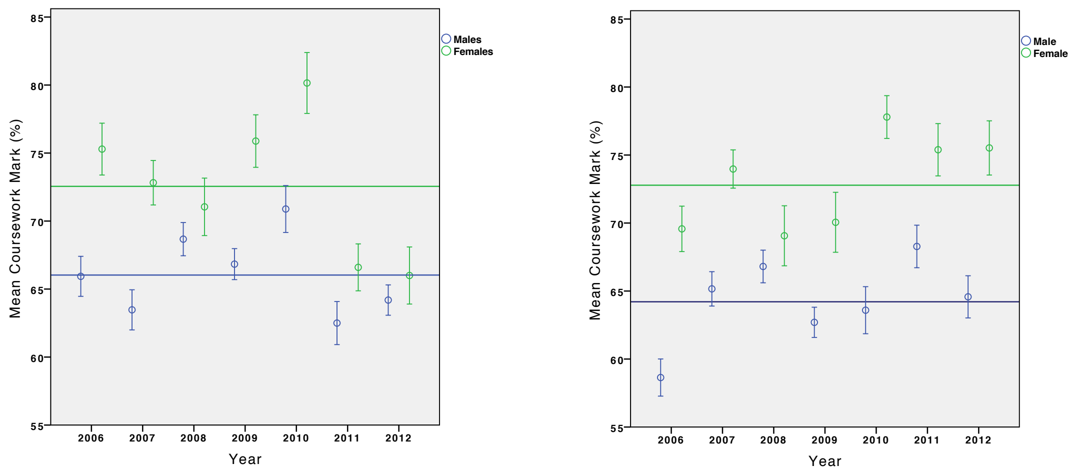
'Physics 1A' and 'Physics 1B' are the two introductory physics courses taken by students in semester one and semester two of their first year of the undergraduate degree. 'Physics 1A' focuses on classical physics of kinematics, dynamics, an introduction to relativity and forces and fields. Students are assessed through both weekly coursework assignments and an end-of-course examination. Whilst the content of the course remained relativity constant between 2006-13, there were some changes in teaching methods and course assessments. During these years the coursework element of the course contributed between 30-33% of the final course mark. This consisted of weekly assignments made up of three physics problems, chosen from those which the students work on in the past week's tutorial. Students took an in-class midterm exam which contributed the equivalent of one weekly assignment. In the 2010-11 academic year the student generated assessment tool PeerWise was first introduced as part of the coursework assessment [95]. Using PeerWise students created their own physics problems online and answered and commented on fellow students' problems. The presentation of first year lectures also changed to a 'flipped-classroom' format in 2011-12, in which students received the content to be covered prior to the lecture and the lectures were used to target areas found difficult by the students through discussion and peer discussion 'clicker' questions [63, 96]. Results of responses to clicker questions will be discussed later in this chapter in section 6.8. For 2011-12 and 2012-13 the end-of-course exam was open-book.

'Physics 1B' introduces the concepts of quantum mechanics and wave particle duality. Lectures touch on topics of superconductivity, thermal physics, nuclear physics and the fundamental properties of matter. As in the first semester course, students complete written weekly assignments worth 20% of their course mark. Since 2010-11, students have completed weekly on-line quizzes and a PeerWise exercise worth the equivalent of one weekly hand-in. The coursework mark comprised the best eight out of ten weekly assignments along with the PeerWise assessment (10%) and weekly online quizzes (10%). They also begin laboratory experiments which contribute 20% to their final course mark, with the end-of-course exam contributing 60%.

6.2.2 First year physics coursework results

In 'Physics 1A', at the end of each week of the semester, students were given three course questions from the past week's tutorials to submit for a weekly assignment. Students had previously been given the opportunity to work through these questions in a group environment during the weekly three hour workshop. During this time students could speak to postgraduate tutors about any questions they may have about the exercises. Prior to the release of the first weekly assignment, first year students were given guidelines on how to approach and set out solutions to physics problems. The importance of describing their problem solving procedure was emphasised. Other aspects of what makes a good written answer which were highlighted were the use of force-body diagrams, dimensional analysis and stating explicitly which physics principles apply in the question.

In this, and all results in this chapter, only students who had a recorded non-zero mark for both coursework and the examination were included in the analysis. Although the questions assigned in the weekly homework remained relatively constant during the time period of this study, the number of questions varied over the seven years. In later years some weekly written assignments were substituted by PeerWise exercises, thereby reducing the total number of coursework hand-ins. In order to better compare student performance between year groups, coursework data presented in this section have been modified to exclude all PeerWise, midterm and FCI scores. The coursework marks therefore constitute only the mean weekly written assignment scores.



(a) (b)

Figure 6.1: Mean coursework scores for male and female students in (a) 'Physics 1A' and (b) 'Physics 1B' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

Figure 6.1 (a) illustrates the mean coursework marks for male and female students in the first year 'Physics 1A' course as a function of the academic year. It can be seen that there was some variation in the average coursework mark over the seven years analysed. In particular 2011-12 and 2012-13 showed a decrease in average coursework score compared to proceeding year groups, particularly for females. This observed drop may be attributed to changes in the marking scheme used to assess individual course questions. Each physics problem in a weekly assignment was given a score in four key areas: Strategic Approach, Physics Explanation, Mathematical Execution and Final Answer. Each of these areas was scored out of 5. A score of 4 in each of these categories indicated a correct answer with all appropriate explanations included in the student's answer. In order to achieve a score of 5 students had to demonstrate a particularly insightful physics explanation, efficient approach, evaluation of mathematical correctness (such as limiting cases or dimensional analysis) or evaluation of the final answer itself. Prior to 2011-12, each question in the weekly assignment was scored out of 4 without the additional coding for the fifth 'bonus' mark.

The key result of this investigation is that in each of the seven years female students had a higher coursework mark than male students. Averaged over these seven years, males had a mean score of 66(1)%, compared to females who had a mean score of 73(1)%. Looking at the performance on each individual year, it was found that the gender difference was statistically significant at the 95% confidence level in four out of the seven years, as measured by an independent t-test. The difference between male and female students' coursework marks was not statistically significant in 2008-09 ($p=0.336$), 2011-12 ($p=0.083$) or 2012-13 ($p=0.474$).

The majority of students who complete the 'Physics 1A' course continue onto the second semester 'Physics 1B' course. Looking at the final written assignment results for 'Physics 1B' shown in Figure 6.1 (b), it was found that, as in the first semester course, female students had a higher average coursework percentage score (73(1)%) than male students (64(1)%) over the seven years. There was a large variation in coursework marks over this time period. Females significantly outperformed male students in each of the seven years, with the exception of 2008 ($p=0.357$).

6.2.3 First year physics lab results

In addition to weekly written coursework assignments, in the second semester 'Physics 1B' course students were assessed on four lab experiments over the course of the 11 weeks. This was students' first exposure to practical physics experimentation, as well as their first time completing a lab book. These lab experiments contributed a total of 20% to the final course score. A gender analysis of the average percentage laboratory marks, shown in Figure 6.2, indicated that females had a higher average score in all year groups. Included in this analysis were all students who were recorded as having a non-zero lab score at the end of the semester. This difference was significant in six of the seven academic years. The gender gap was not statistically significant in 2008-09 (p=0.559).

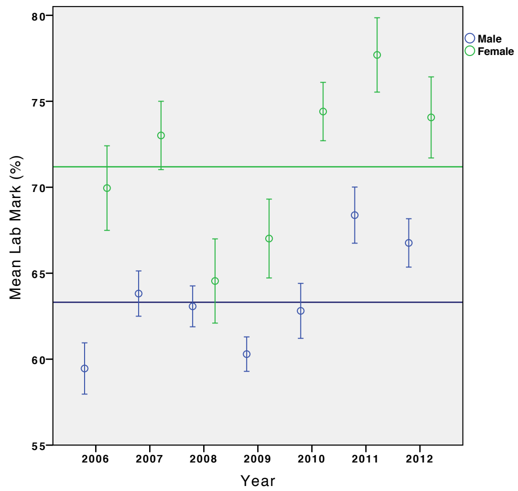
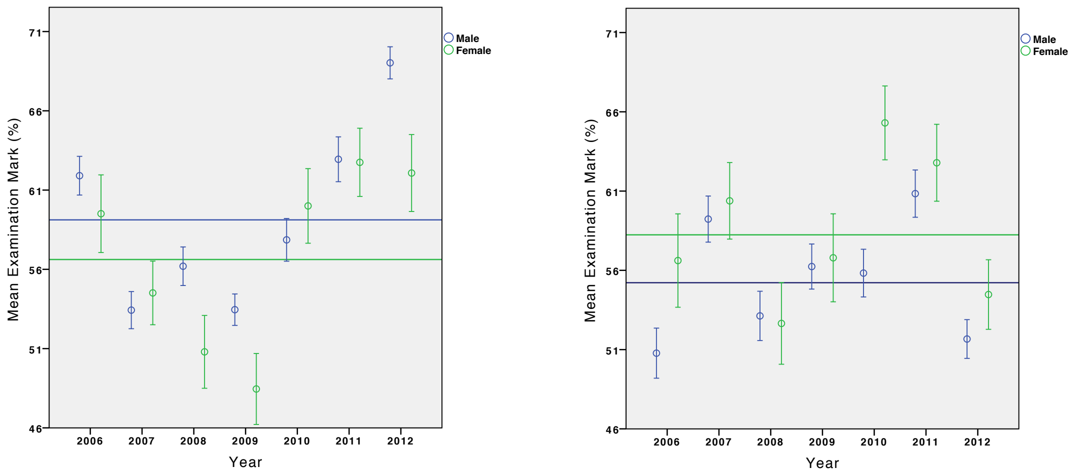


Figure 6.2: Mean lab scores for male and female students in 'Physics 1B' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average lab mark for males (blue) and females (green) over the seven years.

Averaged over these seven years, females had a lab mark of 71(1)%, while males had an average of 63(1)% ($p<0.001$). Each laboratory experiment is assessed through the completion of a lab book which students can work on out of lab hours. In this respect, because students have the ability to discuss the experiment and their results with their peers and complete their lab book outside of contact hours, it is perhaps not completely unsurprising that these results were in agreement with those seen in first year coursework, where females consistently outperformed males.

6.2.4 First year physics examination results

A similar analysis was conducted on final examination scores for 'Physics 1A', the results of which are shown in Figure 6.3 (a). The end-of-course examination comprised a series of compulsory short answer questions and the choice of long answer questions, of which students must complete two. Male students had a significantly higher average percentage examination score (59(0.2)%) than female students (57(1)%) across the seven years examined (p=0.012). Despite this significant difference overall, males did not outperform females in all years and there was large variation in the mean examination scores depending on academic year. In both 2007-08 and 2010-11, female cohorts had a higher average examination score than males, although this difference was not found to be significant at the 95% confidence level (p=0.651 in 2007-08 and p=0.414 in 2010-11). In 2011-12 both genders had effectively equal scores in the end of semester examination. The gender gap was significantly different in 2008-09 (p=0.032), 2009-10 (p=0.034) and 2012-13 (p=0.004).



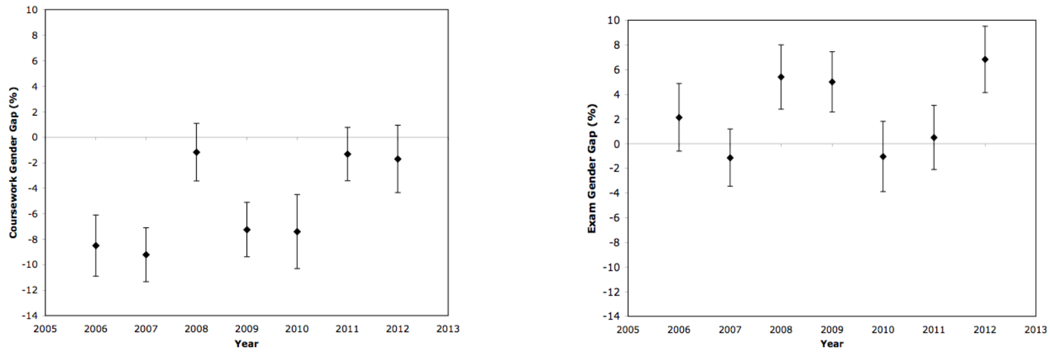
(a) (b)

Figure 6.3: Mean examination scores for male and female students in (a) 'Physics 1A' and (b) 'Physics 1B' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

Data collected from the second semester 'Physics 1B' course, shown in Figure 6.3 (b), also indicated large variation in the gender gap and mean examination scores depending on the year in question. In this case, averaged over the seven semesters, females had a higher mean examination score (58(1)%) compared to males (55(1)%) (p=0.010). Although female students had a slightly higher average percentage examination score in six of the seven years, this gender gap was not statistically significant except in 2010-11. In this year, females performed significantly better than their male counterparts (p=0.001).

6.2.5 Comparison of coursework and examination performance in first year

As in Chapter 3, the gender gap (G) was defined as the average male score minus the average female score. A positive gender gap indicates that male students performed more highly than female students. Similarly, a negative gender gap indicates that females performed better than males.



(a) (b)

Figure 6.4: Average percentage gender gap (G) between male and female (a) coursework and (b) examination scores in 'Physics 1A' for each academic year between 2006-13. Error bars represent the standard error.

Figure 6.4 (a) illustrates that in all years the coursework gender gap for the 'Physics 1A' course was negative. The average coursework gender gap between 2006-13 was -5.2%. Looking at the gender gap for 'Physics 1A' examination scores the gender gap was positive in five of the seven years (Figure 6.4 (b)). The magnitude of the average gender gap for examination scores (2.5%) was also much smaller than for coursework, suggesting that male and female students' performances are perhaps more comparable in exam scenarios. It was not the case that the years that demonstrated the largest negative gender gap in coursework saw the largest gender gap in examination performance.

The relationship between students' attainment on coursework and examinations was examined by first binning student results into quartiles of approximately equal size based on their overall coursework marks. The mean examination scores were then calculated for each gender in each of these four quartiles. A single factor ANOVA test was carried out to determine if any statistically significant differences existed between the coursework marks collected over seven consecutive years. Differences did exist for mean coursework marks between year groups. As a result, all seven sets of 'Physics 1A' data could not be legitimately combined into one data set. Presented below are the results of the combined 2011-12 and 2012-13 data, shown in Figure 6.5. There existed no statistical difference between the mean coursework marks for the whole class nor for the male or female cohorts in these two years. (Results from courses for each year between 2006 and 2010 can be found in Appendix H (Figures H.1 - H.5). In both years the course and lectures followed the same format and weekly coursework assignments were graded using the same marking rubric.

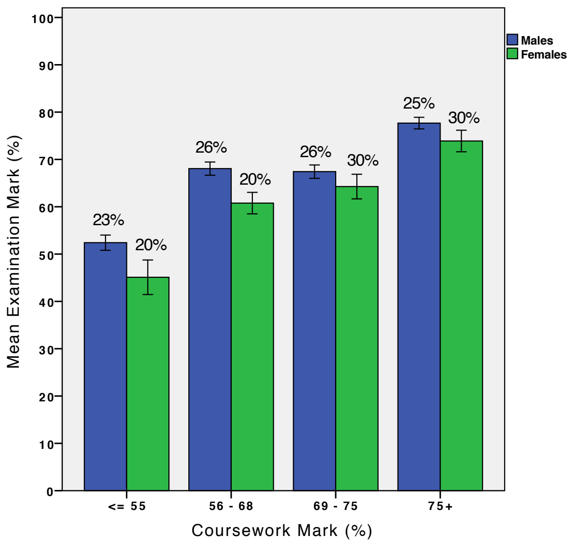


Figure 6.5: End-of-course examination scores of 'Physics 1A' as a function of male and female coursework performance quartiles for the combined 2011-13 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=364 and N(females)=105.

For 2011-13, females had a lower mean examination score in each of these four quartiles than males. Those female students who were in the same quartile as males nevertheless underperformed in the examination compared to their male peers. No statistical difference existed between the examination performance of males and females who were in the lowest coursework performance quartile. On average, males who scored less than 55% (the upper limit of the lowest quartile) on coursework had a mean examination score of 52(2)%, compared to corresponding females who had a mean score of 45(4)%. Similarly, the performance difference was not significant between genders in the top two coursework quartiles. The mean end-of-course examination scores for the second lowest quartile did however show a statistical discrepancy between genders in final examination results (p=0.014).

As in the quartile comparison for the FCI pre-test scores shown in Chapter 3, the male population was distributed evenly across the four coursework performance quartiles for the 2011-13 data set. Approximately 25% of the male cohort existed in each quartile. The distribution of the female population showed that a higher proportion of female than male students existed in the upper two quartiles. This skew in female distribution towards top scores on weekly written assessments was also particularly apparent in 2006-07, 2007-08 and 2009-10, in which less than 10% of the entire female population were found in the lowest coursework performance quartile (See Appendix I). This trend is consistent with the fact that we see an overall female bias in coursework performance. This result was very different to that seen in the FCI, in which almost half the female population was in the lowest performance quartile, reiterating the idea that, because the coursework assignments contain additional concepts to those tests in the FCI, male and female students' overall semester coursework marks may differ to the post-test FCI trends.

6.3 Second year physics performance

Up until this point the exploration of the gender gap in undergraduate physics at the University of Edinburgh has focused on differences in our first year cohort, with the exception of the FMCE and BEMA tests, discussed in Chapter 3, which showed no gender discrepancies. In order to determine whether the observed gender differences in first year were intrinsic to the course, performance on courses in other years of the degree programme was investigated between 2006-13.

'Physics 2A' is a second year undergraduate course taken in the first semester. Almost all students enrolled on this course are physics majors, with only a handful of students taking this as an elective or as a requirement for another degree course such as geophysics. This course aims to provide an introduction to several key topics which form the basis of future physics degree courses: special relativity, electromagnetism, optics and classical dynamics. In addition to lectures and tutorials, students complete a weekly three hour module on Java programming and data analysis. The end-of-course examination contributes 70% to the final course mark. The weekly assignments contribute 10%, with computing and data analysis contributing 20%.

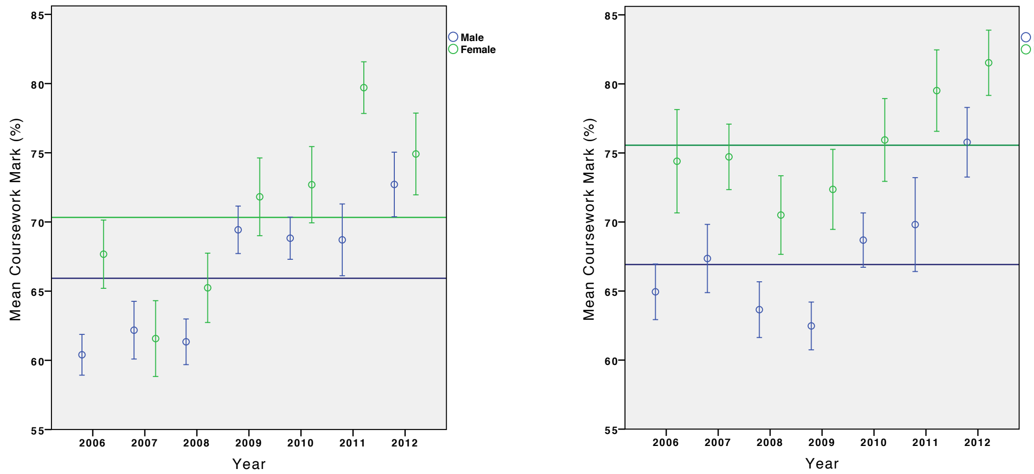
'Physics 2B' is a second semester course taken in the second year of the undergraduate physics programme. Lectures focus on the dynamics of waves (sound, electromagnetic and mechanical), as well as concepts of interference and diffraction. Students are also given an introduction to quantum theory and thermodynamics. Once again, the final examination contributes 70% of the final course mark, with an additional 15% coming from weekly assignments. Students spend three contact hours a week completing lab experiments worth 15% of their final mark.

In 2012-13 there was a restructuring of the second year physics degree courses. The 'Physics 2A' and 'Physics 2B' courses were replaced by the first semester 'Classical and Modern Physics' course and the second semester 'Physics of Fields and Matter' course. 'Classical and Modern Physics' is designed to introduce pre-honours physics students to dynamics, waves, special relativity and quantum physics and is assessed 20% by coursework and 80% by the end-of-course examination. 'Physics of Fields and Matter' focuses on electromagnetism and condensed matter physics. Although not completely analogous to data collected from academic years between 2006-12, results from 2012-13 for both courses have been included in analysis for comparison of second year gender behaviour.

6.3.1 Second year physics coursework results

In both 'Physics 2A' and 'Physics 2B' students completed weekly coursework assignments consisting of physics problems taken from the previous week's tutorial sheet. The coursework scores compared in this section refer only to the weekly hand-in assignments and do not include the data analysis, computational or practical components of the courses.

Average coursework marks for male and female students enrolled on 'Physics 2A' (and 'Classical and Modern Physics' in 2012-13) are shown in Figure 6.6 (a). In six out of seven years female students outperformed male students in the weekly coursework assignments. This was not the case in 2007-08, where male (62(2)%) and female students (62(3)%) had equal average coursework marks. Although females displayed higher average scores in six year groups, this difference was only statistically significant in 2011-12 (p=0.005). Over the seven years, females had an average coursework score of 70(1)%, compared to males who had an



(a) (b)

Figure 6.6: Mean coursework scores for male and female students in (a) 'Physics 2A'/'Classical and Modern Physics' and (b) 'Physics 2B'/'Physics of Fields and Matter' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

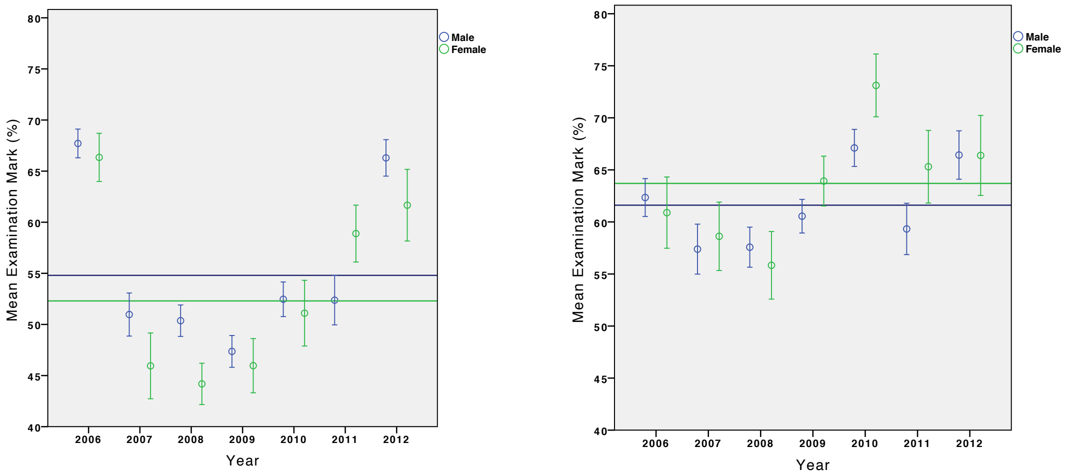
average score of 66(1)% ($p<0.001$). There were large variations between years, with average coursework scores showing a general upwards trend.

The gender difference was more pronounced in the second semester course ('Physics 2B'/'Physics of Fields and Matter'). As shown in Figure 6.6 (b), female students once again performed more highly than male students in each of the years examined. In four academic years this gender gap was statistically significant at the 95% level. This was not the case in 2008-09 ($p=0.054$), 2010-11 ($p=0.057$) or 2012-13 ($p=0.190$).

6.3.2 Second year physics examination results

Final examination results for the core second year physics courses were also analysed for differences in male and female performance. Mean scores for each academic year of the 'Physics 2A'/'Classical and Modern Physics' course are shown in Figure 6.7 (a). Although males had a marginally higher average examination score (55(1)%) over the seven years compared to females (52(1)%), there was large variation from year to year. As was the case in first year physics courses, the examination gender gap was much smaller than for coursework. In fact, only the 2008-09 population showed a statistically significant gender difference (p=0.027).

Similar trends were seen in the second semester 'Physics 2B'/'Physics of Fields and Matter' results shown in Figure 6.7 (b). Both male and female students showed very similar mean examination scores in each of the academic years, with no year group showing statistically different gender performance. Both courses demonstrated a large range in final examination scores depending on the academic year analysed. These fluctuations were not entirely unexpected due to the changes that occur in examination questions from year to year, as well as changes to both the course lecturing staff and curriculum in 2012-13.



(a) (b)

Figure 6.7: Mean examination scores for male and female students in (a) 'Physics 2A'/'Classical and Modern Physics' and (b) 'Physics 2B'/'Physics of Fields and Matter' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

6.4 Summary of physics results

Results from both first and second year data sets indicated that, although there was consistent evidence to support the hypothesis that females perform better than males in coursework and laboratory assessments, results from end-of-course examinations showed no clear gender trends. It should be noted that, while the content of the weekly coursework assignments remained relatively constant over the past few years, examination questions changed on a yearly basis.

Research by the University of Colorado analysed gender gaps in students' physics examination and coursework performance over seven semesters of teaching [3]. In each of these semesters they found that female students scored consistently higher than male students on coursework. This, along with results from the University of Edinburgh discussed in this chapter, offer evidence to support the theory that variations in the learning or assessment environment may encourage different learning gains in different genders [65]. Kost *et al* also noted that across all semesters males showed consistently higher examination marks than females [3]. This was not a trend reflected in results from our pre-honours physics courses. Although males students did outperform females in some year groups, this was not universally the case, nor were these gender differences always statistically significant. This may suggest that both genders perform more equally on examination style assessments and they in turn may not favour male learning.

Looking at overall 'Physics 1A' course scores for each academic year, which take into account both examination and continual assessment contributions, there existed no statistically significant gender gaps, with the exception of results from the 2012-13 year group in which males had a higher overall course score compared to females ($p=0.029$). For the second semester 'Physics 1B' course there were no statistically significant gender differences in final course marks in four of the seven years. Females had a significantly higher overall course mark in 2006-07 ($p=0.004$), 2010-11 ($p<0.001$) and 2012-13 ($p=0.028$). In the second year 'Physics 2A' and 'Physics 2B' courses, six of the seven academic years analysed showed no difference for males and females. The exception to this was the 2011-12 year group for 'Physics 2A', where females had a slightly higher final course mark than males ($p=0.032$), and the 2010-11 year group for 'Physics 2B', where females once again outperformed males ($p=0.022$).

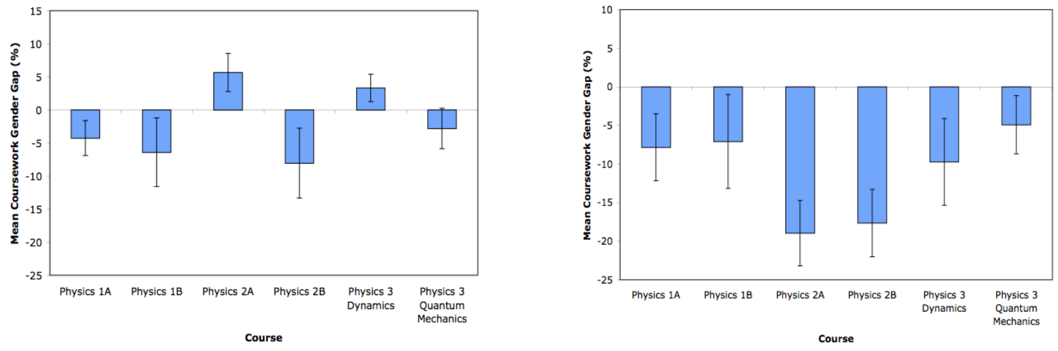
6.5 Longitudinal study of gender performance in physics

All results presented in this chapter so far have explored the consistency of differences between male and female performance in individual courses over several consecutive academic years. There existed changes in course structure, lecturing staff and incoming student selection. These can all have an effect on measured class performance. As discussed in section 6.1, by undertaking a fully longitudinal study, the same students can be tracked through the entirety of their degree program and the progression of the gender gap explored. This analysis can help to establish whether the observed gender gap in first year persists and if it is a characteristic of the cohort, or whether certain courses or years of the degree programme have different effects on the gender assessment gap.

Two year groups of students were followed from their first year through to their fourth year of study; those who commenced their studies in 2006-07 and those who began in 2007-08. In the results presented in this section, data was collected from core courses in each of the four years of the BSc degree programme and a fully longitudinal study carried out (section 6.1). Seven core physics courses were analysed. These included the first year ('Physics 1A' and 'Physics 1B') and second year ('Physics 2A' and 'Physics 2B') courses discussed earlier in this chapter. Two third year courses, 'Dynamics and Relativity' and 'Quantum Mechanics', were also included, as well as the 'Condensed Matter Physics' course taken by students in the fourth year of their degree. Only students for whom there existed both coursework and end-of-course examination results for each course were included in the analysis. In total 20 male and 13 female students fulfilled these requirements in 2006-07 and 31 male and 7 female students were included in 2007-08.

6.5.1 Coursework gender gaps

Looking first at the students starting their university degree in 2006-07, Figure 6.8 (a) depicts the mean percentage gender gap in total coursework mark for each of the core degree courses analysed between 2006 and 2010. The negative gender gap in the first year courses indicated that female students had a higher average coursework score than male students, as was reported in section 6.2.2. Following these thirty-three students into their second year of study showed a positive performance gender gap in the 'Physics 2A' coursework. Interestingly, this again became negative in the second semester 'Physics 2B' course. By the junior honours year of their degree the size of the coursework gender gap was reduced and the sign or direction of this gap once again changed depending on the course: positive in 'Dynamics and Relativity' and negative in 'Quantum Mechanics'. The average coursework gender gap over these three years of the degree programme was -2%. All fourth year senior honours courses contain no assessed coursework, but are assessed solely by an end-of-course examination and are therefore not included in the coursework analysis.



(a) (b)

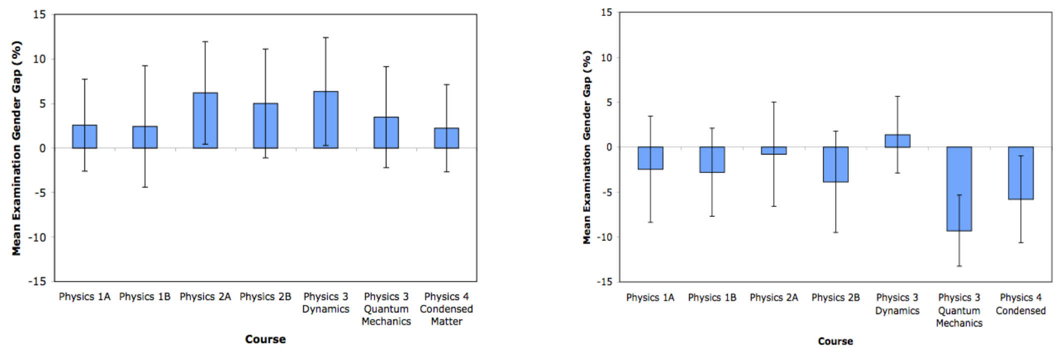
Figure 6.8: Average percentage coursework gender gap for core degree courses for (a) 2006-10 and (b) 2007-11 longitudinal cohorts. Error bars represent the standard error on the mean.

Repeating the longitudinal analysis for the 2007-08 cohort, the coursework gender gap was calculated for the same six courses (Figure 6.8 (b)). The gender gap was negative in each case, indicating that females consistently outperformed males in each course, irrespective of the year of the degree programme or course content. Averaged over their four years of study, the mean coursework gender gap was -11%, a much greater gender discrepancy than that of the proceeding population. There were, however, variations in the magnitude of the gender gaps from course to course. Unlike in the 2006-07 longitudinal analysis, the coursework gender gap was strongly negative for both second year courses. In 'Physics 2A' the female population had a 19% higher average coursework score than males. This was comparable to the second semester second year 'Physics 2B' course where the gender gap was -18%. Once again, the 'Dynamics and Relativity' and 'Quantum Mechanics' courses in junior honours showed a narrower gender gap for this population. Although the size of these gender gaps, particularly in second year courses, are concerning, the magnitude of these large gender gaps must be treated with some caution since the female population size was only seven students.

6.5.2 Examination gender gaps

Results from a longitudinal investigation of the end-of-course examination scores for these two student populations are shown in Figures 6.9 (a) and (b). For the 2006-07 cohort, in each course analysed there was a positive gender gap in the end-of-course examination scores, indicating that, on average, male students outperformed female students. The average percentage gender gap was 4%. However, there were substantial error bars associated which each data point due to the large range in individuals' results and the small number of students in each gender population. As in the coursework results, the second year courses showed the largest discrepancy between male and female students' scores.

The overall trend in the 2007-08 population was slightly different. In six out of seven of the courses analysed the gender gap was negative, although once again there were substantial standard errors on the mean due to small number statistics, particularly for the female cohort. The average gender gap across the four years was calculated to be -3%. In some instances, for example 'Physics 2A' and 'Dynamics and Relativity', the gender gap was minimal.



(a) (b)

Figure 6.9: Average percentage end-of-course examination gender gap for core degree courses for (a) 2006-10 and (b) 2007-11 longitudinal cohorts. Error bars represent the standard error on the mean.

6.5.3 Summary

It is perhaps not unexpected that there exist some disparity between the performance of the two longitudinal cohorts. One of the limitations of this study was the very low number of male and female students included in the analysis.

The students included in the longitudinal study who began their degree in 2006-07 comprised only 20 males (61%) and 13 females (39%). In 2007-08 the number of females was considerably lower, with only 7 female students (18%) completing the coursework and examination elements in all four years of the programme. In addition to the attrition of students across the degree programme, the number of students fulfilling the requirements to be included in such a project may also be limited by students who repeat academic years or enter straight into second year through the direct entry programme and are therefore not included in the analysis.

The majority of the courses taken by both populations indicated a negative gender gap in coursework throughout the degree programme. The exceptions to this were the second year 'Physics 2A' course and the third year 'Quantum Mechanics' course for the 2006-07 cohort. The small population sizes and large statistical errors associated with the data points limit the conclusions that can be made from this study. Although there existed trends in the data to support the fact that in the majority of cases females perform better than males on continual assessments, the size of this gender gap fluctuated greatly both between populations and amongst years of the degree programme. Similarly, calculated examination gender gaps were very small in comparison to their standard errors, suggesting that in some cases male and female students' performance on examinations is relatively balanced. As previously indicated by the first and second year examination results, the measured difference between male and female scores on end-of-course examinations was considerably lower than for coursework, suggesting that this type of assessment may show less gender discrepancy. This lack of gender bias in examination results at the University of Edinburgh does not support results from previously published literature which showed males consistently outperforming females on examinations [3].

6.6 Gender performance in STEM subjects

Physics has a dramatically low proportion of females enrolling on degree courses compared to many other STEM subjects [167]. Having seen a consistent gender gap in coursework performance for undergraduate physics courses at the University of Edinburgh, it was queried whether there existed similar trends in other science degree programmes within the College of Science and Engineering. In the following section the demographics and student performance in core first year chemistry and biology courses at the University of Edinburgh will be presented with an emphasis on the difference in attainment of male and female students. Both disciplines have higher percentages of female students taking introductory courses than physics. As in the physics analysis discussed in the previous section, only students who completed both the coursework and examination requirements for the course were included in the results.

6.6.1 Undergraduate attainment in chemistry

Chemistry students complete two core courses in their first year of study: 'Chemistry 1A' and 'Chemistry 1B'. 'Chemistry 1A' provides an introduction to key topics including chemical bonding, atomic structures and thermodynamics. Students complete a mixture of laboratory and tutorial work. 'Chemistry 1B' follows on directly from the first semester course and focuses on teaching the methods of spectroscopic analysis and chemical reactions. In both courses the end-of-course examination contributes 55% to students' final course mark, with written coursework assignments contributing 20% and a practical examination a further 25%.

Both 'Chemistry 1A' and 'Chemistry 1B' have higher proportions of female students enrolled compared to undergraduate physics courses. There were, however, fluctuations in the number of male and female students included in results for each academic year, as is shown in Tables 6.1 and 6.2. For the last three years the population has remained relatively stable, with females making up just under 50% of the cohort.

6.6.2 First year chemistry coursework results

Figure 6.10 (a) depicts the average coursework marks for male and female students in 'Chemistry 1A' between 2006-13. Although there was considerable variation in the average class coursework mark over the seven year period, it was seen that female students outperformed males in each academic year. This gender gap was statistically significant at the 95% confidence level for all years. The average female coursework score over this time period was 81(1)%. This was

Table 6.1: Number and percentage of male and female students included in each of the analysed years of 'Chemistry 1A' between 2006-13.

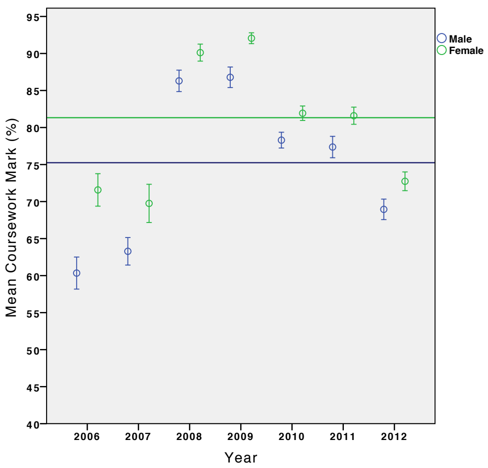
Year	N	N(male)	% Males	N(female)	% Females
2006-07	162	106	65%	56	35%
2007-08	187	135	72%	52	28%
2008-09	254	150	59%	104	41%
2009-10	224	144	64%	80	36%
2010-11	188	95	51%	93	49%
2011-12	180	94	52%	86	48%
2012-13	176	94	53%	82	47%

Table 6.2: Number and percentage of male and female students included in each of the analysed years of 'Chemistry 1B' between 2006-13.

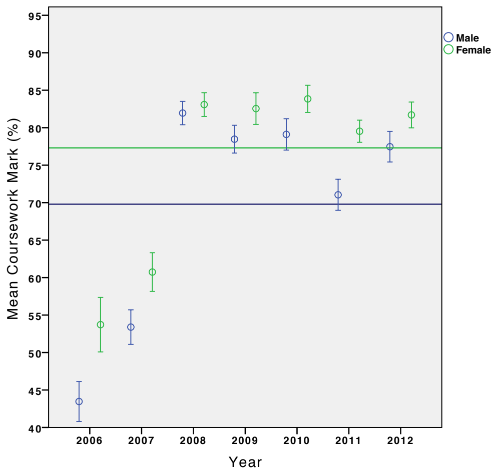
Year	N	N(male)	% Males	N(female)	% Females
2006-07	136	86	63%	50	37%
2007-08	163	117	72%	46	28%
2008-09	221	130	59%	91	41%
2009-10	189	122	65%	67	35%
2010-11	169	85	50%	84	50%
2011-12	165	87	53%	78	47%
2012-13	162	84	52%	78	48%

significantly higher than the male average score of 75(1)% ($p<0.001$). The second semester 'Chemistry 1B' course showed a similar pattern (Figure 6.10 (b)), with females averaging 77(1)% on coursework between 2006-13, compared to males who averaged 70(1)% ($p<0.001$).

In both data sets a sharp increase in average coursework marks was noted after 2007. There existed several changes to these courses which may account for this. The first was a change in the minimum entry requirements. The average entry qualifications have shifted upwards considerably over the last few years (from BBBC for Scottish Highers to ABBB, with students typically entering with straight As). Secondly, a weekly on-line submission was introduced to replace a weekly written exercise. Both the written exercise and on-line submission contributed to the final coursework marks, but it was found that grades were slightly higher for the on-line exercise as there was a shift towards marks for



(a)



(b)

Figure 6.10: Mean coursework scores for male and female students in (a) 'Chemistry 1A' and (b) 'Chemistry 1B' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

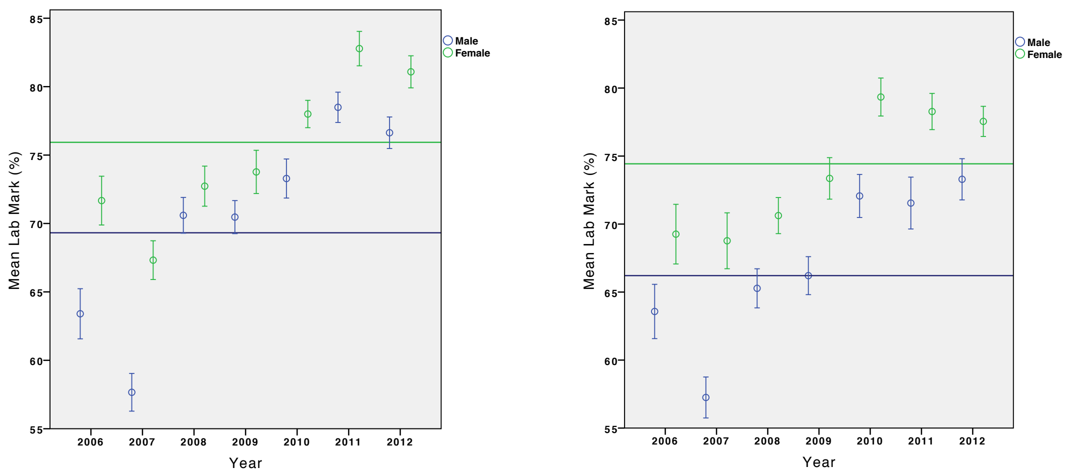
engagement rather than assessment. There were also changes in instructors during this time. All these factors may have had the effect of elevating students' continual assessment scores.

6.6.3 First year chemistry lab results

Analysis of the laboratory scores for 'Chemistry 1A', shown in Figure 6.11 (a), showed that first year female students had higher average marks compared to males in each academic year, a result consistent with that of the first year physics labs in section 6.2.3. In five out of seven years this gender gap was statistically significant. The two exceptions were in 2008-09 ($p=0.285$) and 2009-10 ($p=0.101$). Averaging over the seven years, males had a mean laboratory score of 69(1)% and females a mean of 76(1)%.

The second semester 'Chemistry 1B' course showed a similar trend (Figure 6.11). Once again females showed higher levels of laboratory performance across the seven years. Males had an average of 66(1)% and females had an average of 74(1)%. The difference between male and female practical performance was statistically different in all academic years excluding 2006-07 ($p=0.071$).

In the labs themselves there was a steady increase in mean scores for both genders. There have been a few revisions of the content of the chemistry practicals



(a) (b)

Figure 6.11: Mean lab scores for male and female students in (a) 'Chemistry 1A' and (b) 'Chemistry 1B' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

and the introduction of on-line pre-lab exercises that may have had a bearing on that particular assessment and thus resulted in higher overall marks.

6.6.4 First year chemistry examination results

The format of the end-of-course examinations for introductory chemistry courses consisted of a paper structured as six questions, each divided into a compulsory and optional section. Students were required to complete all six compulsory parts and a choice of any four optional sections in a 2.5 hour time period. This meant that students were compelled to answer questions on all areas of the course content.

Results from first year examinations demonstrated a different pattern to that which was seen in the introductory physics courses. Looking first at 'Chemistry 1A', females outperformed males in the majority of years, with the exception of 2012-13 in which male and female students had almost equal mean examination scores. The gender difference was statistically significant in three academic years (2006-07 $p<0.001$, 2010-11 $p=0.001$ and 2011-12 $p=0.031$). Averaged over the seven year period, males had a significantly lower mean examination score (62(1)%) than females (66(1)%), with a p value of <0.001.

A gender analysis of the second semester 'Chemistry 1B' course also illustrated

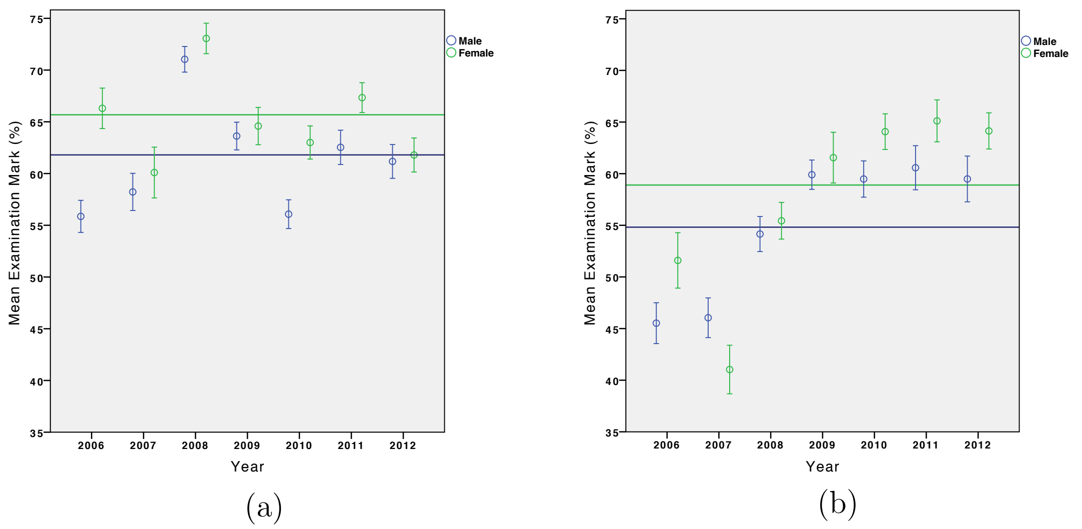


Figure 6.12: Mean examination scores for male and female students in (a) 'Chemistry 1A' and (b) 'Chemistry 1B' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

a consistent gender discrepancy in favour of females. In almost all year groups males had a lower end-of-course examination score, although the difference was not found to be statistically significant in any year. This was not the case in 2007-08, where first year male students taking the course had a higher mean score than females (46(2)% and 41(1)% for males and females respectively). Averaged over all years, females significantly ($p<0.001$) outperformed males in exams, with females having an average score of 59(1)% compared to 55(1)% for males.

6.6.5 Undergraduate attainment in biology

Biology students undertake two core courses in their first year of study. 'Origin and Diversity of Life 1' discusses the structural form and metabolic adaptations of living organisms. It introduces students to genetics and the evolution of species. The second semester 'Molecules, Genes and Cells 1' course focuses on the structure and experimental investigation of cells and nucleic acids. It acts as an introduction to the concepts of biological membranes and cell growth. The end-of-course examination contributes 60% of the final course mark for both these courses. Students are also assessed through coursework (25%) and a practical examination (15%).

The proportion of female students undertaking these two introductory biology courses is relatively high compared to other STEM subjects. Unlike physics, which had an average of 24% female students, biology had approximately 60-65% females in the first year courses investigated here (Tables 6.3 and 6.4).

Table 6.3: Number and percentage of male and female students included in each of the analysed years of 'Origin and Diversity of Life 1' between 2006-13.

Year	N	N(male)	% Males	N(female)	% Females
2006-07	442	187	42%	255	58%
2007-08	481	195	41%	286	59%
2008-09	419	157	37%	262	63%
2009-10	508	229	45%	279	55%
2010-11	370	134	36%	236	64%
2011-12	369	123	33%	246	66%
2012-13	382	129	34%	253	66%

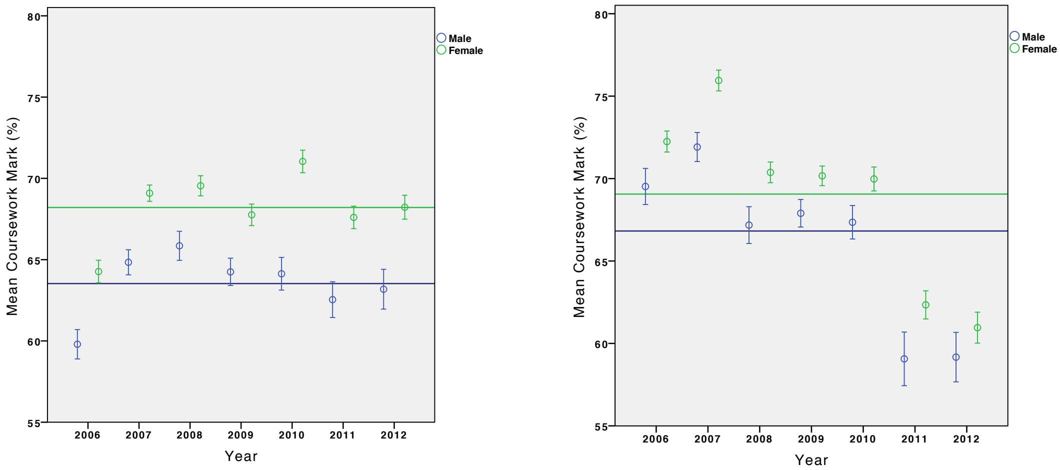
Table 6.4: Number and percentage of male and female students included in each of the analysed years of 'Molecules, Genes and Cells 1' between 2006-13.

Year	N	N(male)	% Males	N(female)	% Females
2006-07	363	147	40%	216	60%
2007-08	406	163	40%	243	60%
2008-09	354	121	34%	233	66%
2009-10	429	195	45%	234	55%
2010-11	316	109	34%	207	66%
2011-12	307	94	31%	213	69%
2012-13	314	106	34%	208	66%

6.6.6 Biology coursework results

Mean coursework marks for the first semester 'Origin and Diversity of Life 1' course are shown in Figure 6.13 (a). For each academic year the calculated mean coursework score for female students was consistently higher than that for male students. Statistical tests found the gender differences to be significant at the 95% confidence level in all seven years investigated. Over this seven year period, male students had an average coursework score of 64(0.3)% and female students had an average of 68(0.4)% ($p<0.001$).

A similar trend was seen in the second semester 'Molecules, Genes and Cells 1' course. Once again, in all years, female students outperformed males in continual assessments. This difference was statistically significant in all years between 2006 and 2011. Although a consistent gender gap was seen in the other two year groups, this difference was not significant (for 2011-12 p=0.077 and for 2012-13 p=0.293). Averaged over the seven year period, females had a mean coursework mark of 69(0.3)% compared to males who had a mean mark of 67(0.2)% ($p<0.001$).



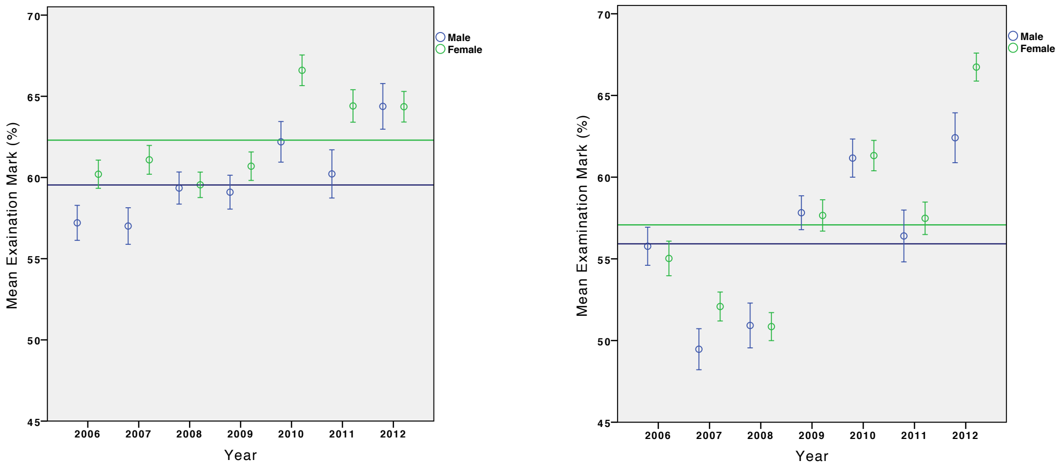
(a) (b)

Figure 6.13: Mean coursework scores for male and female students in (a) 'Origin and Diversity of Life 1' and (b) 'Molecules, Genes and Cells 1' for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

6.6.7 Biology examination results

Gender differences were dramatically lower in examination results, as can be see by Figures 6.14 (a) and (b). Although, in 'Origins and Diversity of Life 1', females had statistically higher scores in four of the academic years, some academic years, such as 2008-09 and 2012-13, showed equal performance levels. When averaged over all the academic years investigated male and female students had statistically different ($p<0.001$) overall scores (60(0.3)% for males and 62(0.3)% for females).

The 'Molecules, Genes and Cells 1' course showed considerable variation in final examination score, most likely due to changes in the examination questions from year to year. Averaged over the seven years, females had a mean score of 57(0.2)% and males a mean score of 56(1)%. There existed very little difference between genders on examinations on this course (p=0.314). In fact, in four year groups male and female populations had equal examination results. The gender gap was significant in 2012-13 (p=0.014). For both courses the gender difference is much less pronounced than in coursework results, as was the case in both chemistry and physics first year undergraduate courses.



(a) (b)

Figure 6.14: Mean examination scores for male and female students in (a) ‘Origin and Diversity of Life 1’ and (b) ‘Molecules, Genes and Cells 1’ for each academic year between 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework mark for males (blue) and females (green) over the seven years.

6.6.8 Overall course marks for first year chemistry and biology courses

By comparing overall course marks for first year undergraduate chemistry and biology courses it was noted that gender gaps existed between cohorts in several academic years. For ‘Chemistry 1A’, female students had a statistically higher final course mark than male students in three years (p<0.001 in 2006-07, p=0.001 in 2010-11 and p=0.015 in 2011-12). The difference between male and female performance was slightly more pronounced in the second semester ‘Chemistry 1B’ course, in which females once again produced significantly higher course marks than males in four of the seven year groups (p=0.022 in 2006-07, p=0.012 in 2010-11, p=0.017 in 2011-12 and p=0.047 in 2012-13).

In the first semester introductory biology course, female students statistically outperformed males in all but two academic years. In 2008-09 (p=0.127) and

2012-13 (p=0.149) there was no difference in male and female performance at a 95% confidence level. For the second semester biology course the only year that showed a statistical significance between gender performance was 2007-08 (p=0.008) in which male students had a lower overall course mark compared to females.

6.7 Comparison of gender gaps across three STEM disciplines

In this chapter gender differences in student performance on course assessments in three STEM subjects have been analysed. As previously mentioned, the populations of these three disciplines have very different gender profiles, with the proportion of female students enrolled on first year courses dramatically lower in physics. This provides the opportunity to compare the magnitude of gender gaps in coursework and examination results with the level of female participation, in order to test the hypothesis that having a more balanced gender population decreases the performance discrepancy.

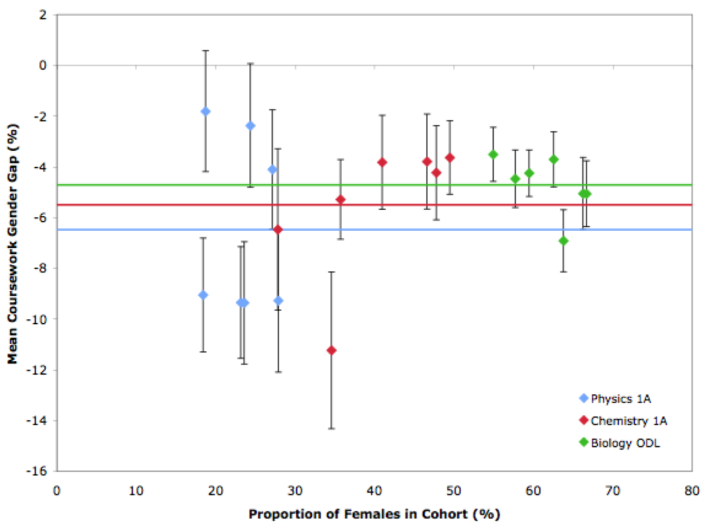


Figure 6.15: Mean coursework gender gap for first year undergraduate courses 'Physics 1A', 'Chemistry 1A' and Biology 'Origin and Diversity of Life 1' as a function of the proportion of female students in each year's cohort for 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average coursework gender gap for 'Physics 1A' (blue), 'Chemistry 1A' (red) and Biology 'Origin and Diversity of Life 1' (green).

Figure 6.15 depicts the mean percentage coursework gender gap for first semester first year physics, chemistry and biology courses as a function of the proportion of females in each year's cohort between 2006-13. In each case, the gender gap was negative, indicating females had a higher overall coursework score than males across all three disciplines. The average coursework gender gap was calculated for each of the courses. There existed a gender gap of -6.5% for physics, -5.5% for chemistry and -4.7% for biology. Although this may initially insinuate that, as the proportion of females in the population increases (or the gender proportions become more balanced), the performance gap decreases, there were large variations in the average gender gap from year to year and large error bars associated with each data point. This suggests that cohort demographics are not the only factor affecting course performance. Course content and instructor influences may also have an effect, both of which had variations over the course of the time period examined.

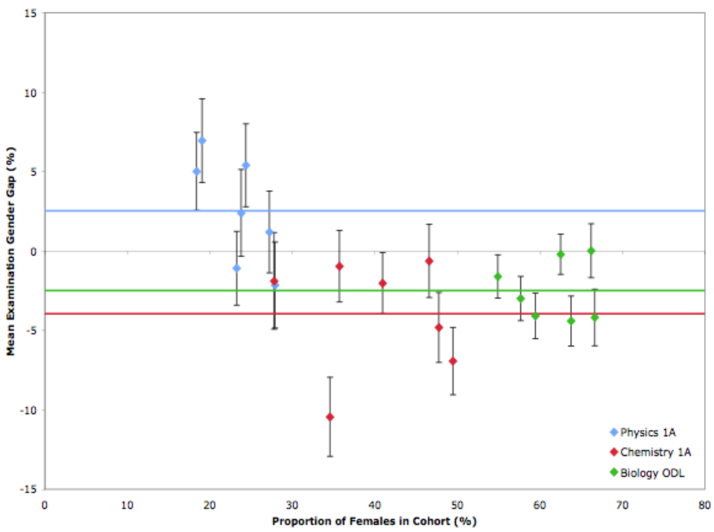


Figure 6.16: Mean examination gender gap for first year undergraduate courses 'Physics 1A', 'Chemistry 1A' and Biology 'Origin and Diversity of Life 1' as a function of the proportion of female students in each year's cohort for 2006-13. Error bars represent the standard error on the mean. Horizontal lines represent the average examination gender gap for Physics 1A (blue), Chemistry 1A (red) and Biology Origin and Diversity of Life 1 (green).

The same analysis was conducted on the end-of-course examination results, the results of which are shown in Figure 6.16. Once again there were large variations in average percentage gender gap for each course depending on the year group. As previously noted, the mean gender gap for examinations in each course across the seven years was smaller than that for coursework results. For chemistry and biology this gender gap was negative, indicating that, on average, female students had a higher overall examination score. For chemistry the average gender gap was -4.0% and for biology the gender gap was -2.49%. The average gender gap for physics students in the first semester was positive, indicating a higher average performance by male students (2.5%). Overall, there appeared to be no direct correlation between the proportion of females in the cohort and the resulting examination gender gap, although physics was more likely to have a gender gap in favour of males.

6.8 Formative assessment in first year physics

In addition to continual weekly written assignments and an end-of-course examination, peer discussion in-lecture questions have been recently introduced into the 'Physics 1A' course. The use of peer discussion and interactive engagement techniques to actively involve students has been reported to show an increase in overall course performance in different subject areas [50, 51, 59]. As noted in Chapter 1, Peer Instruction (PI) has the advantage of altering the lecture environment to encourage more two-way discussion between the instructor and the students, as well as inviting students to engage in discussion amongst each other. Previous studies have shown that the act of explaining a concept to another peer can improve a student's own understanding of the topic [60]. There is also a benefit to the students to whom it is being explained.

One of the primary ways in which PI has been implemented in courses, including our introductory physics classes, is through the use of personal response systems, more commonly referred to as 'clickers' [170]. During a lecture, the instructor poses a question to the class and asks them to use their clicker to vote individually on what they believe is the correct answer from a series of multiple-choice options (or a true and false answer). After this, students are given a few minutes to discuss the question and possible solutions with their neighbours. The class then re-votes and the lecturer initiates discussion based on students' responses. There has been some deliberation in studies as to whether instructors should display class results after the first round of voting, before students begin discussing possible answers with their nearest neighbours. A study by Perez *et al* investigated whether displaying these preliminary responses resulted in biased students' answers in the second round of voting [59]. They found that participants who were shown a bar chart of the distribution of initial responses were 30% more likely to change their response to the most popular option. This effect was more pronounced for a true or false question than for a multiple-choice question. In all of the PI episodes discussed in this chapter the initial class responses were not revealed to the student prior to the second round of voting.

Peer Instruction clicker questions have been used extensively in 'Physics 1A' lectures, which have followed an inverted classroom structure since 2011-12 [63, 172]. Students were provided with a clicker for the duration of the course, enabling them to be assigned an electronic ID for post-lecture analysis. It was emphasised to students that no marks were given for correct or incorrect answers, nor was any credit given for participation. The lecturer followed the PI method outlined above, allowing students approximately one and a half minutes before pre-discussion answers to the multiple choice questions were collected via the electronic clickers. Where the proportion of correct responses was initially recorded as being between 30-70% of the class, students were explicitly encouraged to discuss the question and their reasoning for choosing their answer with other students in the lecture theatre. In this respect, not all clicker questions were suitable for productive peer discussion. If the proportion of students answering correctly is too low or too high, the probability of finding someone with a different viewpoint for effective discussion is low. Students were encouraged to move around the lecture theatre if their surrounding neighbours shared the same initial viewpoint, thereby inviting them to discuss it with someone who voted differently to themselves, although in practice students discussed the questions with those sitting around them. A second round of individual voting then took place, followed by a discussion led by the instructor.

6.8.1 Results of Peer Instruction in 'Physics 1A'

The use of PI clicker questions in 'Physics 1A' lectures in 2011-12 provided a large source of data, with 41 PI episodes[1] occurring over the course of the 11-

[1] A PI episode refers to the administration of the same clicker questions twice during a single lecture, with students participating in peer discussion between two rounds of voting.

week semester. These PI episodes accounted for approximately half of all clicker questions used during the course. The concepts of Newtonian mechanics dealt with in the FCI were covered in the first five weeks of instruction. A total of 65 clicker questions were asked in this section of the course, of which 14 involved full PI discussion (pre-vote, group discussion, revote). These 14 questions will be discussed in further detail in this section.

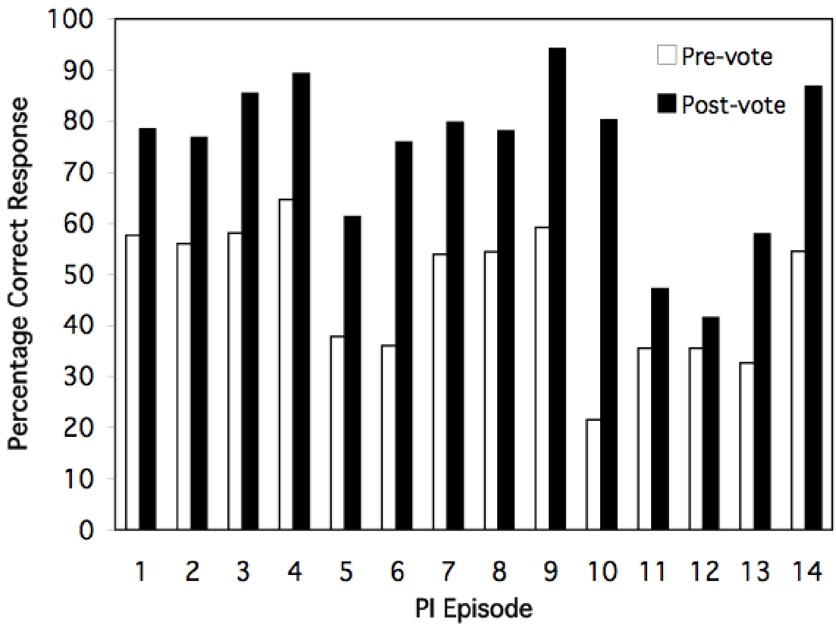


Figure 6.17: Percentage correct responses in the pre- and post-vote for PI episodes.

The participation levels for each question ranged from 56% to 77% of the total students enrolled on the course. The fluctuation in participation rates depends on a number of factors: student attendance in lectures, whether students have brought their personal clicker handset to the lecture and, finally, student participation in the voting process. There existed very little difference in the proportion of the male and female cohorts who responded for the majority of questions. In all results presented below only matched student data was included (students entering responses for both pre- and post-discussion vote). Those students who participated in only one round of voting were eliminated from analysis.

Looking first at the percentage of correct responses for the class as a whole, shown in Figure 6.17, it is apparent that each PI episode showed an increase in student learning from the pre-vote to the post-vote. The extent of this improvement varied considerably depending on the individual question, with learning gains ranging from 0.09 on PI episode 12 to 0.86 on PI episode 9.

6.8.2 Gender performance on in-class PI clicker questions

It has been shown earlier in this chapter that there exist trends in the gender attainment gap depending on assessment type. In this section the difference in male and female performance on PI episodes as a whole will be compared, along with gender performance on individual questions which showed interesting gender discrepancies.

Analysis of the 14 full PI question pairs showed that males slightly outperformed females in the average pre-discussion percentage of correct responses over the first five weeks, although this difference showed no statistical significance at the 95% level. Figure 6.18 depicts the range of percentage of correct responses for these 14 in-lecture clicker questions. Once again, both genders had large learning gains between pre- and post-discussion responses. Female students showed a smaller range in percentage of correct responses despite both genders having similar mean pre- and post-vote percentages of correct answers.

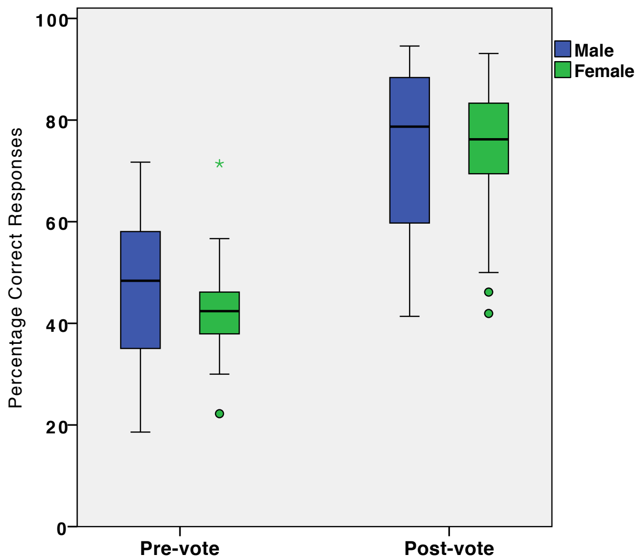


Figure 6.18: Boxplot of average pre-vote and post-vote percentage correct responses to Newtonian mechanics PI questions as a function of gender in 2011-12.

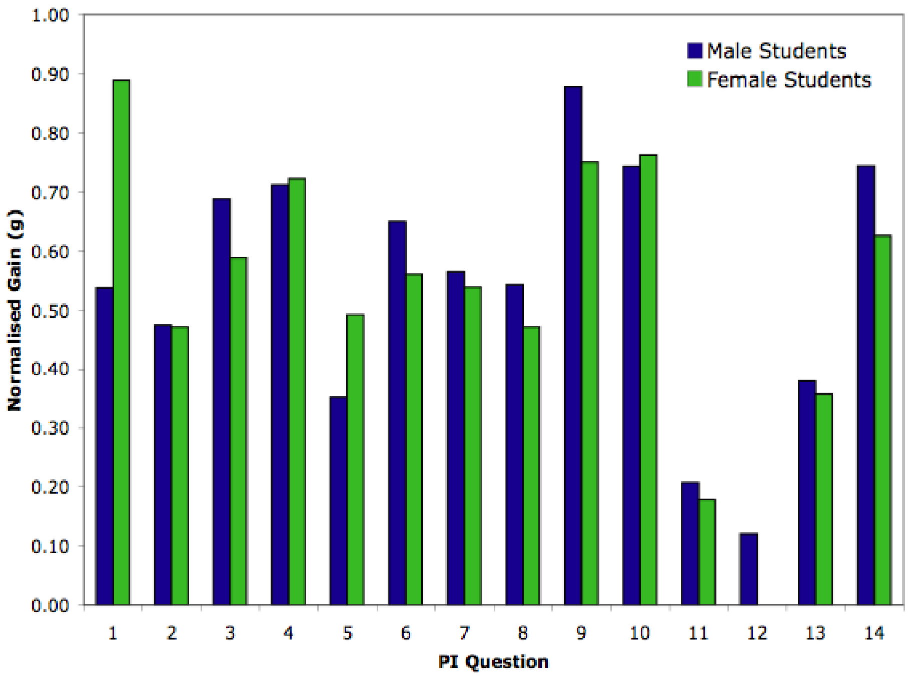


Figure 6.19: Average normalised gains for 14 PI 'clicker' questions as a function of gender for 2011-12.

Despite there being no significant gender difference in students' overall performance on PI questions, individual questions showed signs of gender differences in performance, with males having a much higher percentage correct rate compared to females on some questions and a greater percentage of females answering correctly compared to males on others. Figure 6.19 shows the normalised gain for the 14 analysed PI episodes as a function of gender. The average normalised gain varied considerably from question to question, giving an indication of unsuccessful and successful PI episodes. This may not be unexpected as these PI episodes placed Newtonian concepts of force and motion in a wide range of contexts. Female students had a higher normalised gain than males on four of the PI questions analysed. In this section the responses to three PI episodes, selected to demonstrate different gender gaps, will be discussed in more detail.

Question 1

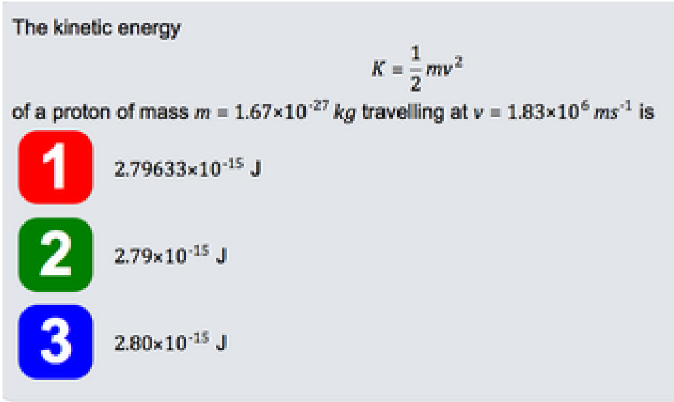
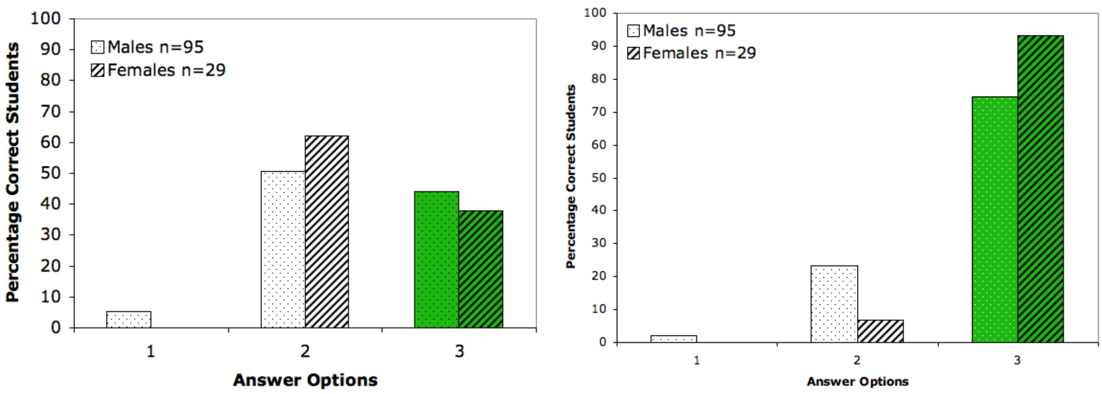


Figure 6.20: PI episode 1: 'Rounding Correctly'.

The first PI episode in 'Physics 1A' was not a conceptual question, but instead asked students to consider how to round their final answers to a sensible number of significant figures. A copy of this question is shown in Figure 6.20. It was nevertheless a good example of a successful PI episode. Analysis of the whole class performance on the pre- and post-discussion votes (Figures 6.21 (a) and (b)) showed that the most popular incorrect answer in the first round of voting was answer option 2. Following discussions with fellow students, the majority of students chose the correct response in the post-vote (option 3).



(a) (b)

Figure 6.21: 'Rounding Correctly': (a) Pre-vote and (b) post-vote responses to PI episode 1 as a function of gender. The correct answer was option 3 (highlighted in green).

Both genders showed very low numbers of students answering correctly on the pre-vote. Initially, only 38% of females correctly chose option 3. This was 6% lower than the proportion of males who had answered correctly in the pre-discussion vote. Interestingly, after peer discussion, females dramatically outperformed males. Percentage correct scores for females exceeded 93% (a normalised gain of 0.89), compared to 74% for males (a normalised gain of 0.54).

Another way to explore the effect of peer discussion amongst students is to look at the transitions of students' answers, as was done in section 4.3. When a student answers the same question twice their responses can be categorised into four different types of transitions; right-to-right, right-to-wrong, wrong-to-right and wrong-to-wrong. As mentioned previously in Chapter 4, in order for the student population to record an increase in learning gain, the number of wrong-to-right transitions must exceed the number of right-to-wrong transitions. The high normalised gain witnessed for female students is reflected in a higher percentage of students (55%) changing their response between the pre- and post-votes for females than for males (36%). Of these transitions, no female students and only 3 male students made a right-to-wrong transition.

Question 9

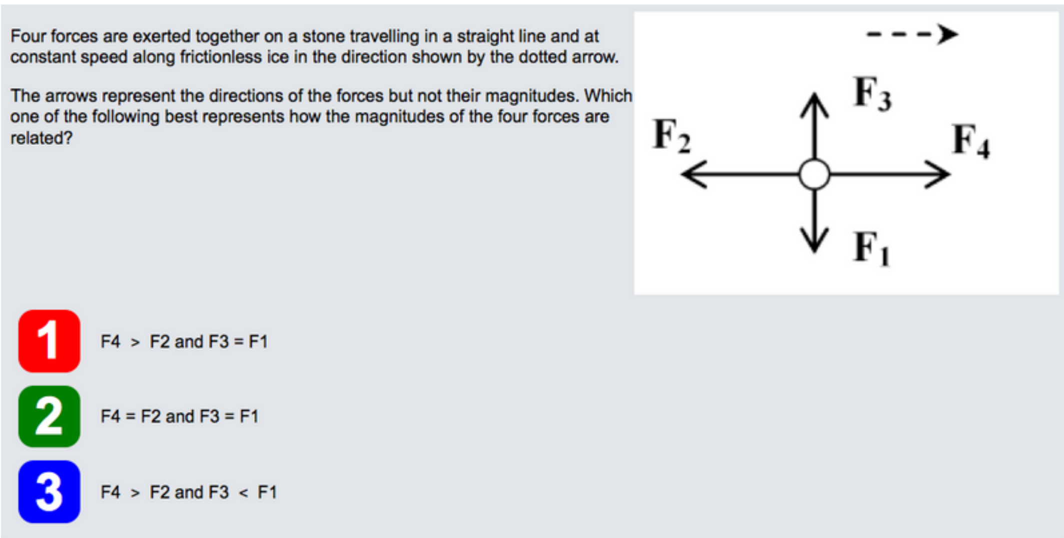
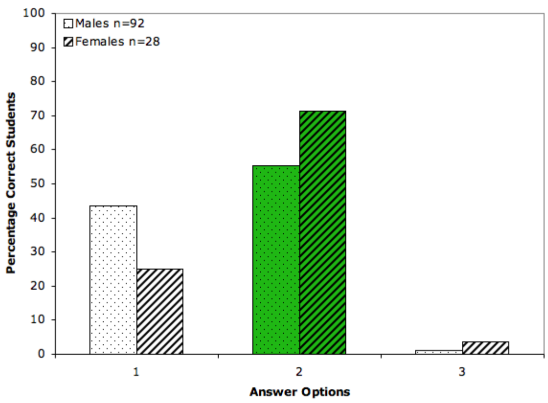
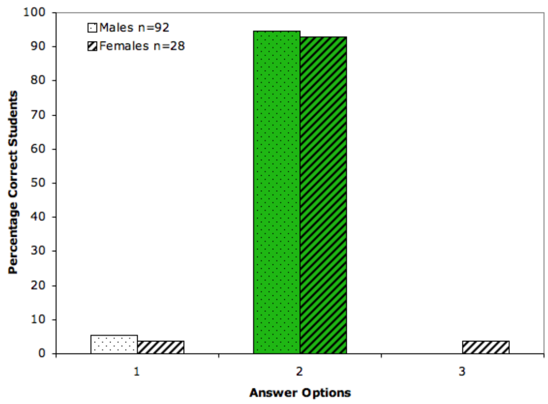


Figure 6.22: PI episode 9: 'Four Forces'.

PI episode 9, shown in Figure 6.22, is another example of a successful PI episode. Respondents are asked to choose which option best describes the relative magnitudes of four forces acting on an object travelling in a straight line at a constant speed on a frictionless surface in the indicated direction. Pre-vote responses for the whole cohort showed that, while 59% of the class correctly identified the correct answer (option 2), a further 39% of students believed the forces would be unbalanced, despite the stone traveling at a constant speed. After having been given time to discuss the question amongst themselves, without any further instruction from the lecturer, the second round of voting showed a dramatic increase in correct responses, with 94% of students answering correctly.



(a)



(b)

Figure 6.23: 'Four Forces': (a) Pre-vote and (b) post-vote responses to PI episode 9 as a function of gender. The correct answer was option 2 (highlighted in green).

Looking at the difference between male and female responses in Figure 6.23 (a), it was found that a much higher percentage of females than males chose the correct answer in the pre-vote, with 71% of females and 58% of males choosing the correct response. Male students did however have a higher overall normalised gain (0.88) than females (0.75). After peer discussion both populations showed very high levels of understanding (Figure 6.23 (b)). In the post-vote 95% of males and 93% of females answered correctly. A total of 41% of male students changed their response between votes, only 1 of the 38 students changing from right-to-wrong. In comparison, only 21% of females changed their response, reflecting the lower learning gain for females. All of these females had a wrong-to-right transition.

Question 12

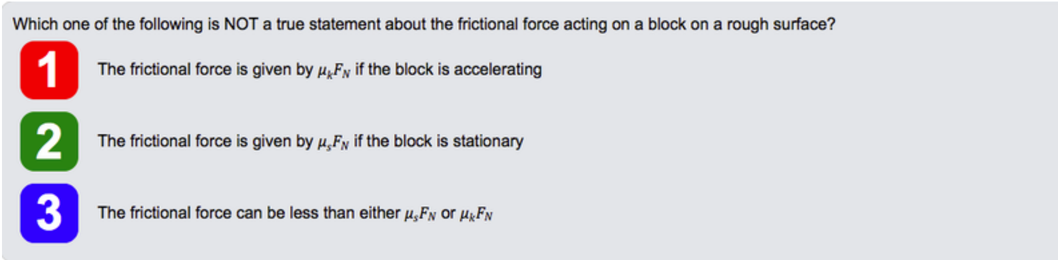
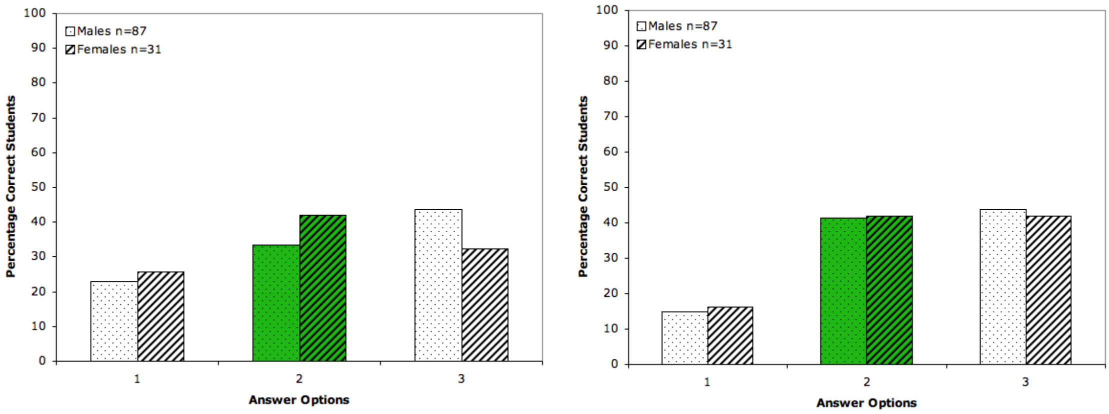


Figure 6.24: PI episode 12: 'Friction true or false'.

Peer discussion was less effective for PI episode 12. This question, shown in Figure 6.24, asks students to consider what they know about how the force of friction is determined for an object on a rough surface, a topic which caused a large amount of confusion for students on the FCI qualitative interviews discussed in Chapter 5. Students would have been introduced to this topic in the pre-reading completed before the lecture. Looking at pre-vote responses, only 36% of the cohort correctly answered option 2 (the frictional force is given by $\mu_s F_N$ is the block is stationary). Post-test results showed very little improvement, perhaps suggesting that peer discussion amongst students was ineffective, maybe due to a lack of conceptual knowledge on which effective discussion could be based. In this case it was required that the lecturer engage the students in discussion of the topic of friction after the second round of voting.



(a) (b)

Figure 6.25: 'Friction true or false': (a) Pre-vote and (b) post-vote responses to PI episode 12 as a function of gender. The correct answer was option 2 (highlighted in green).

This question is an example of an unsuccessful PI episode in which both genders had minimal normalised gains. Males had an average normalised gain of 0.12 and females had a normalised gain of zero. In this case, there was very little evidence overall of students changing their answer between the vote and revote (Figure 6.25). Only 8 out of 31 female students changed their vote after PI discussion, half of whom went from right-to-wrong and half went from wrong-to-right. A higher proportion of the male population changed their answer between voting session (35 of the 87 male students). Of these, 14 students went from right-to-wrong.

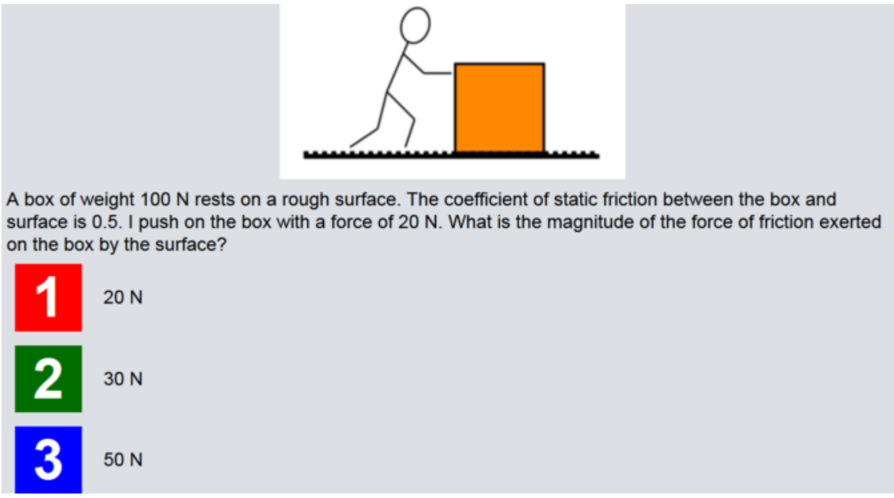


Figure 6.26: Revised PI question on friction.

Another explanation of the reason for students' poor performance is the composition of the question itself. The original PI question was a 'negative' question, meaning that students had to identify which statement was incorrect. Analysis undertaken as part of another project at the University of Edinburgh exploring student conversations during peer discussion suggested that the wording of the question may have contributed to students' confusion [173]. Recordings of students' conversations drew attention to their confusion over the symbols used and found that this activated a formula-based approach. As a result, the question was subsequently revised in the following year. A copy of the revised question can be seen in Figure 6.26. This question places friction in a real world context and asks students to consider the forces acting on the box before it begins moving. Although the two questions on friction are presented very differently they are both intended to focus on the same fundamental physics concept. When this question was used in lectures in the following academic year students showed a significantly higher normalised learning gain (0.51).

6.8.3 Discussion of PI results

Overall, results suggest that there is an equally beneficial effect on student learning and engagement in lectures for both male and female students. The average normalised gain for both genders on PI episodes in the first five weeks of 'Physics 1A' was comparable to those seen overall on the FCI and are consistent with those found by Hake for interactive courses [44]. Large variations were seen in performance on individual PI episodes, with some questions showing higher learning gains than others. As was noted in the three PI episodes discussed above, male and female students did not necessarily show equal learning gains after peer discussion, suggesting that there may be a gender element to performance on a question by question basis, although it is difficult to conclude from the relatively small sample of PI episodes analysed during the Physics 1A course whether these gender differences are more pronounced for questions relating to a specific physics concept. It may be difficult for students to engage in constructive peer discussion if too large a proportion of students lack understanding of the physics concept being tested. In such cases, it might be necessary for the instructor to provide additional information to the class as a whole. Results have also indicated that the wording and presentation of the question can affect students' understanding and subsequent performance.

A study by Smith *et al* at the University of Colorado explored whether the resulting increase in number of correct responses from students arose as a result of an increase in understanding, or whether it was a consequence of the influence of higher performing students on their neighbours in lectures [58]. They discovered that groups of students showed learning gains even if there was noone in the group who knew the correct answer in the pre-vote. It is therefore necessary to consider that some results may be a consequence of peer influences or even correct answers resulting from incorrect reasoning [174]. The instructor may consequently perceive these correct answers as students having a clear understanding of the topic. It might therefore be unreasonable to assume that all increases in measured normalised gains in our first year course are the direct result of improved conceptual understanding. The influence of the dialogue between students during a PI episode on students' cognitive reasoning has been explored in a study at the University of Edinburgh with first year undergraduate physics students [175]. Gaining information about what students discuss with peers during in-lecture PI questions can help guide the lecturer's post vote discussion and explanations, as well as inform any improvements in the creation of future PI questions.

6.9 Chapter discussion and summary

We have seen in this chapter that recorded performance levels have shown measurable gender differences depending on the assessment method employed: coursework, laboratory assessments, examinations and Peer Instruction. In this chapter the physics performance of undergraduate students has been considered, first through the comparison of performance of consecutive year groups on the same course and secondly through a fully longitudinal study over the entirety of a cohort's degree.

In terms of coursework attainment, physics students at the University of Edinburgh have shown consistent gender differences. First year undergraduate female students demonstrated higher coursework marks than male students in almost all cases, and in many instances this gender gap was statistically significant. Although slightly less pronounced in the first semester of second year, female students once again had an higher average total coursework mark over seven years of analysis. Practical lab work completed by students continued this trend, with male students consistently under-performing compared to female students. The similarity between results for weekly assessments and experimental marks may not be entirely unexpected, since both allow students the opportunity to study the problems in a collaborative environment (in groups during tutorials and with their lab partners in physics practicals) before handing in individual assignments. These results were also in agreement with those reported in the literature [3, 56]. The relatively high coursework performance of females compared to males has been noted across many science and social science disciplines [56, 66, 68]. In a study of chemical engineering students, team-based cooperative learning was highly valued by both genders but particularly women

[166]. Felder *et al* noted that male students were more likely to comment that they found it beneficial to explain the problems and working to fellow students, whilst females commented that the greatest benefit of group work was having the content explained to them by others. One possible explanation for the female bias in coursework may be that females are more diligent and spend more time working on their coursework assignments outside of class time. Mau and Lynn reviewed previous studies which look at the amount of homework completed by males and females and noted that in high school females spent more hours doing homework [176]. The correlation between homework and final course mark was also greater for female students than males.

When analysing end-of-course examination results, there appeared less distinction between genders. Despite male students showing a statistically higher average examination score in 'Physics 1A', when averaged over the seven years the gender gap was not consistent from year to year. In two of the academic years investigated female students showed a marginally higher average examination score, whilst males significantly outperformed females in three years. For the second semester 'Physics 1B' course female students had a higher average examination score over the seven years, although once again this gender difference was not significant except in 2010-11. Core second year courses showed very similar mean examination scores for male and female students. In both first and second year the gender performance gap was much smaller for examination scores than coursework scores. Despite consistent gender gaps on continual assessments, overall course marks showed no statistical differences between final scores for male and female students in the majority of cases.

Comparing results from consecutive year groups in core courses indicated large variations in mean assessment scores, depending on the academic year analysed. Longitudinal analysis conducted on two different year groups allowed for the magnitude of the gender gap to be established for the same cohort of students across four years of the degree programme. By keeping the population constant for each course, the progression of the gender gap could be explored and it could be determined whether different courses showed different levels of discrepancy between male and female performance. However, one consequence of this was the very small number of students included in each data set, particularly for females. Different gender gap trends were witnessed for each longitudinal cohort.

For students starting in 2006-07, the negative coursework gender gap seen in the first year of study was positive in the 'Physics 2A' course, before once again becoming negative in the second semester. Third year also showed a change in the direction of the gender gap depending on the course, although the gap had decreased greatly. Students who began their studies in 2007-08 had consistently negative coursework gender gaps, although the size of this gender gap varied. A similar trend was seen in examination results. As in the pseudo-longitudinal study, the gender gaps for end-of-course examinations was much smaller than those for coursework. The small population sizes resulted in large standard errors on the mean scores, making it difficult to gain definitive conclusions from the data.

Perhaps one of the most intriguing outcomes from this research is in the comparison of data from physics undergraduates with that of first year chemistry and biology students. These three STEM subjects have very different proportions of female students in their undergraduate populations, with the number of female physics students dramatically lower than in chemistry or biology. In all three degree subjects the first year coursework gender gap was negative. The mean gender gaps between 2006-13 for the three disciplines tentatively suggested that as the proportion of females in the cohort increases, the difference in gender performance on continual assessment decreases. However, this trend should be treated with some caution. There were large fluctuations in the size of this gap depending on the academic year, suggesting that cohort demographics are unlikely to be the only factor affecting course performance and other factors such as changes in lecturing staff or the presentation of the course content could potentially affect the relative performance of male and female students. Examination results across the three sciences showed that the size of the examination gender gap was much smaller than for continual assessments.

Results from the University of Edinburgh did not show as well defined a trend as at the University of Colorado, where the examination results consistently favoured male students [3]. While males had higher scores than females for almost all years of the 'Physics 1A' course, this was not the case in chemistry and biology where females consistently scored more highly than males.

How these gender differences in individual assessments manifest themselves in the overall course grade is interesting. For physics courses, the majority of year groups showed no performance gap. In those that did, females often outperformed males. This may be a consequence of females significantly outperforming on continual assessment combined with a relatively balanced gender performance on the final exam. First year undergraduate chemistry and biology courses showed greater gender differences in final course marks than introductory physics courses, with female students significantly outperforming males in several academic years.

The strength of PI is seen in results from analysis of in-lecture clicker questions showing large learning gains between pre- and post-vote responses for the majority of questions. This suggests that the introduction of PI has had an overall positive effect on student learning. PI encourages students to engage with the material and practice qualitative reasoning as they discuss potential solutions with their peers. Participation levels suggested no gender bias in participation, although it was impossible to discern from the data how many students attended the lecture, but did not participate in the voting process. Data collected from each 'clicker' question indicated that the number of students answering each question posed by the lecturer fluctuated within a single lecture, suggesting that some students consciously decided to refrain from voting at a particular instance. It can be unclear from purely quantitative data how PI truly affects students' reasoning. Difficulties can arise if students answer correctly but their reasoning is incomplete. They may be confident that they fully understand the topic because they voted correctly. In order to tackle this problem the instructor could reinforce the correct conceptual reasoning after the answer has been revealed. Instructors must be aware that low normalised gains could also be attributed to the wording or the question itself, as was seen in PI episode 12. Editing of PI questions may be necessary as students may approach and interpret the problem in unexpected ways. There did exist some differences in gender performance on individual 'clicker' questions. On some questions females showed a higher normalised gain than males and vice versa. Once again, this may suggest that different genders find different questions more conceptually difficult or find that they contradict their intuitive beliefs.

Chapter 7

Attitudes and Beliefs about Learning Physics

The role played by students' attitudes towards studying science has been widely investigated over the last few decades, with particular emphasis placed on exploring reasons why students choose to pursue scientific careers. It is hoped that by identifying key reasons why students choose to study a subject, or equally why they have chosen not to continue with a subject, a better understanding of the current participation levels can be achieved. One major concern is the gender participation discrepancy in many of the science subjects, in particular physics, in which females are dramatically under-represented.

Evidence has also suggested that students' performance on science courses can be strongly influenced by the way in which they think about the subject [78, 166]. These attitudes can manifest themselves early on in a student's education. There are many areas which contribute to students' 'attitudes to science': motivation to studying science, enjoyment of the subject, attitudes of peers and parents and achievement or fear of failure in the subject [80]. In this thesis the focus is on the attitudes of undergraduate students towards studying physics in a university environment. This is explored using both quantitative and qualitative techniques.

In this chapter, results are discussed from the implementation of the Colorado Learning Attitudes about Science Survey (CLASS) [84] which has been employed to measure student attitudes toward learning and studying science. The change in attitudes of first year undergraduate physics students after one year of study, in particular any difference in these attitudinal changes with respect to students' gender, was explored. The attitudes of these students were compared to an expert response predetermined by the survey instrument. Whilst the expert responses to each of the survey items is independent of the student cohort being assessed, it is unknown if this expert opinion, developed by US physics faculty during the validation of the survey, is consistent across other academics with a different educational background, with no published work available in this area at the time of this PhD. Presented in this thesis are the results of a study examining the attitudes of physics faculty and industry members in the UK. A comparison of these responses with those of the US academics is made as well as a gender analysis of their responses.

In addition to students' attitudes and beliefs measured by the CLASS survey instrument, qualitative data examining students' future intentions, both at the start and completion of their degree programme, will be discussed. The reasons for students choosing to study their degree subject, as well as their intended degree exit point, were explored through the use of surveys and qualitative interviews. Finally, after having completed their undergraduate degree, graduating students were given the opportunity to provide feedback on their degree courses, the key results of which will be presented in this chapter.

7.1 Colorado Learning Attitudes about Science Survey (CLASS)

The Colorado Learning Attitudes about Science Survey (CLASS) is one of the most widely recognised survey instruments used to measure students' attitudes to studying and learning science and has been widely implemented in North America and worldwide [84, 85]. The survey quantifies students' attitudes in comparison with a view predefined by physics academics. A full description of the survey instrument and its design and validation has been presented in Chapter 2 of this thesis. A copy of the survey can be found in Appendix E.

The CLASS survey consists of 42 attitudinal statements marked on a five point Likert scale (from strongly agree to strongly disagree). A percentage favourable score (percentage agreement of the student responses with those of the predefined 'expert' response) and a percentage unfavourable score (percentage to which the student responses are in disagreement with the 'expert' opinion) can be determined. Student scores can be analysed both as an overall favourable or unfavourable score and as scores from groups of questions. The authors of the survey carried out a factor analysis of these statements in order to determine a series of categories into which the statements could be grouped. Eight categories were created: *Personal Interest*, *Real World Connection*, *Problem Solving General*, *Problem Solving Confidence*, *Problem Solving Sophistication*, *Sense Making/Effort*, *Conceptual Understanding* and *Applied Conceptual Understanding*. Of the 42 statements in the final survey, 27 fall into these eight categories. A further 9 statements are included in the overall favourable and unfavourable scores. There are 6 statements which were concluded as being *"not useful in their current form"* and are subsequently not included in the calculation of 'expert-like' thinking scores, but remain in the final survey. Some of the statements in the survey are placed in one or more of the eight categories.

CLASS has previously been employed at the University of Edinburgh to look at students' attitudes prior to 2010 [177]. In this thesis CLASS has been used for the purpose of investigating the change in attitudes of students over their first year in the physics degree programme and to examine potential differences between different demographics, in particular gender, in the way in which students think about studying physics.

7.2 First year undergraduates' attitudes and beliefs towards physics

In this section the attitudes and beliefs of first year undergraduate students at the University of Edinburgh will be discussed. Differences between male and female students' attitudes towards studying physics were investigated as well as changes in each year group over the period of one academic year. In this section the results from three consecutive academic years between 2010 and 2013 will be presented. The CLASS survey was distributed to students in the first year undergraduate physics courses ('Physics 1A' and 'Physics 1B') to determine the extent to which undergraduate students agree with the defined 'expert' view on studying physics and how this agreement changed over the course of their first year of study.

Students at the University of Edinburgh received this survey during a weekly tutorial workshop in their first week of undergraduate tutorials (week 2 of semester 1) and again in the final week of their first year (week 10 of semester 2). At the beginning of the tutorial students were given a short introduction explaining the survey and its use in investigating how students thought about science and their approaches to learning science. It was emphasised that they should respond as honestly as possible about each statement and that it was important that they should complete it individually. The survey was presented in paper form and no course credit was given for its completion. On average students took approximately ten minutes to complete the survey. Only the responses of students who completed both administrations of the survey (pre- and post-instruction) were used in analysis.

Because the CLASS survey was carried out as a pre- and post-instruction survey to explore the change in students' opinions over time, it was not possible to completely anonymise the survey results. Students were asked to provide their student ID numbers on their written questionnaires. Participants were assured that this was required only for analysis of data and that no individuals would be identifiable after matching of pre- and post-instruction data. The collection of this information allowed for the data to be analysed, not only for changes in the whole class population, but also for potential gender differences and differences between major and non-major students. The total number of matched (pre- and post-instruction) surveys is shown in Table 7.1.

Table 7.1: Number of completed matched CLASS surveys for 2010-11, 2011-12 and 2012-13 as a function of gender.

Cohort	2010-11	2011-12	2012-13
All	104	76	123
Male	71	54	104
Female	33	22	19

7.2.1 Results of whole class survey

A statistical ANOVA test carried out on data from the 2010-11, 2011-12 and 2012-13 academic years showed that there was no statistical difference in percentage favourable pre-test scores between the three populations. However, there was a statistical difference at the 95% confidence level between the 2010-11 and 2011-12 cohorts post-instruction scores (p=0.007), with the average favourable post-test score for the 2011-12 cohort (74(2)%) significantly higher than that for 2010-11 (68(2)%). Here, and in all subsequent values, the number in the brackets refers to the standard error on the mean. Because of this statistical difference, the three data sets were not combined into one larger data set. Results of the overall favourable and unfavourable expert-like thinking scores for each academic year will be discussed.

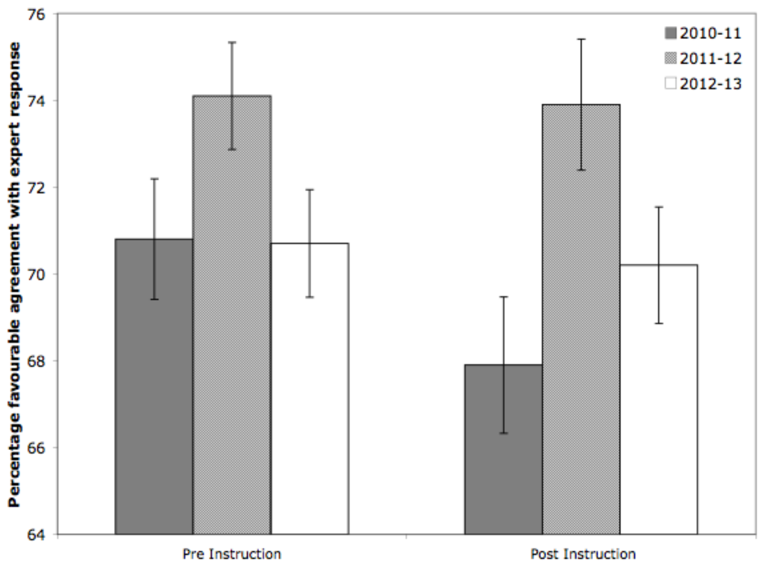


Figure 7.1: Percentage of favourable expert-like thinking pre- and post-instruction for 2010-11, 2011-12 and 2012-13 first year undergraduate students. Error bars represent the standard error on the mean. N(2010-11)=104, N(2011-12)=76 and N(2012-13)=123.

Looking at results for each academic year, as shown in Figure 7.1, it was seen that first year students in each year group entered the undergraduate programme with high levels of expert-like thinking compared to results published by survey authors for American students [84]. Survey authors discussed results collected from a calculus-based first semester physics course at a North American research university which showed an overall percentage favourable score of 65% [84].

The 2010-11 cohort indicated a 71(1)% agreement with the experts' responses at the start of the academic year. After two semesters of teaching, which made use of interactive engagement techniques such as those discussed in previous chapters, the population showed a decline in expert-like thinking, with a post-instruction average of 68(2)%. This drop was found to be statistically significant using a paired t-test (p=0.009). For the 2011-12 year group, students once again began their degree programme with a high level of expert-like thinking, with a 74(1)% favourable agreement with the predefined expert response. At the end of the academic year students were resurveyed, but showed no change in their level of expert-like thinking (p=0.891). A similar pattern was seen in 2012-13. Students who entered with a level of expert-like thinking of 71(1)% were found to have a post-instruction CLASS score of 70(1)% (p=0.689).

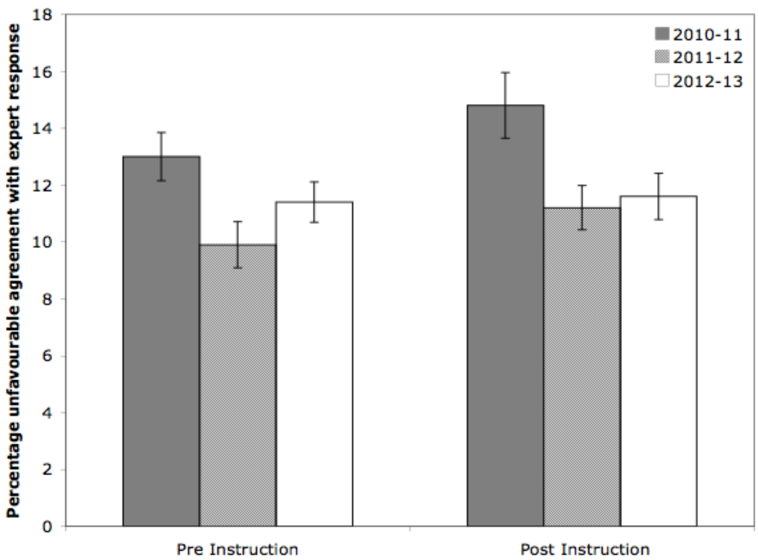


Figure 7.2: Percentage of unfavourable expert-like thinking pre- and post-instruction for 2010-11, 2011-12 and 2012-13 first year undergraduate students. Error bars represent the standard error on the mean. N(2010-11)=104, N(2011-12)=76 and N(2012-13)=123.

As well as the percentage favourable score, which considers the extent to which students agree with the expert responses to each statement, we can look at the percentage unfavourable score, which measures the extent to which students disagree with the expert responses. The results of this analysis can be seen in Figure 7.2. The percentage of favourable and unfavourable responses do not necessarily always add up to 100% due to the fact that respondents are able to chose a neutral response for survey statements. In 2010-11 students commenced the degree programme with an average unfavourable score of 13(1)%. This level of disagreement rose significantly to 15(1)% at the end of the academic year (p=0.038). In 2011-12 students had a pre-test unfavourable score of 10(1)%. After two semesters of teaching this increased very slightly to 11(1)%, although this increase showed no statistical difference (p=0.691) as measured by a paired t-test. In 2012-13 the level of disagreement with the expert opinion once again remained constant between the pre-test (11(1)%) and post-test results (12(1)%). The pre-instruction percentage of unfavourable expert-like thinking for the 2010-11 cohort was significantly higher than in 2011-12 (p=0.009) and this statistically significant difference remained in the post-test (p=0.011). In addition to this, the 2010-11 cohort had a significantly higher post-instruction unfavourable score compared to the 2012-13 cohorts (p=0.025). There were no differences between the measured attitudes of the 2011-12 and 2012-13 year groups. It is difficult to compare the level of disagreement with expert responses found at the University of Edinburgh with results from other institutions because only favourable scores have been reported in published literature.

Previous studies have shown that a consistent drop in expert-like attitudes is witnessed in courses [84, 85], unless the curriculum is specifically designed to address student epistemologies. There have been some cases showing an increase in expert-like thinking of students [88, 89]. Although a statistically significant decrease in expert-like thinking was measured in 2010-11 at the University of Edinburgh, the two following years showed no significant change and therefore differ from results seen in the literature. This was also not in line with previous data found at the University of Edinburgh [177]. The timing of this change from the previously consistent decrease in favourable scores coincides with the switch to an inverted classroom lecture format. Whether this change is a direct result of new teaching methodologies is difficult to conclude, but offers the prospect for future research. If this lack of decrease in expert-like attitudes was to persist in future years' data it may suggest that teaching methodologies, such as Just in Time Teaching and peer discussion in lectures, may help maintain students' favourable attitudes towards studying physics.

7.2.2 Comparison of major and non-major students' expert-like thinking

As discussed in Chapters 2 and 3, not all students enrolled on first year introductory physics courses are intending to study for a physics degree. Students can be classed as 'majors' or 'non-majors' depending on their degree intention. In Chapter 3 large differences were found in major and non-major students' conceptual understanding of Newtonian mechanics, despite both cohorts achieving the same entry qualifications. By examining CLASS scores for these two cohorts, it can be determined whether they also have different attitudes to learning and studying physics. Recalling the fact that both physics majors and non-majors must achieve the necessary entry requirements for enrolling on a physics degree, the only distinguishing feature between the two populations is their degree intention. This in turn may suggest that we should expect a difference between the attitudes towards studying this subject. Table 7.2 shows the number of students, and proportion of each year's cohort, who were classed as majors or non-majors and who completed both the pre-instruction and post-instruction CLASS surveys.

Table 7.2: Number of students and the percentage of each cohort as a function of subject major who completed both pre-instruction and post-instruction CLASS surveys.

Year	Major	Non-Major
2010-11	54 52%	50 48%
2011-12	38 50%	38 50%
2012-13	51 41%	72 59%

Table 7.3 shows the pre- and post-instruction percentages of expert-like thinking for majors and non-majors for each of the three academic years investigated. There are several points of interest that can be seen from these figures. First, in each year there existed no statistically significant difference between major and non-major physics students' pre-instruction scores. Major students did have a higher percentage favourable pre-instruction score compared to non-majors in two of the three years. The difference was particularly apparent

in 2010-11 where majors had a favourable score of 73(2)%, compared to non-majors who had 69(2)% (Figure 7.3 (a)). For this year non-major students also showed a significant drop in expert-like thinking after two semesters (p=0.011). As a result, the post-instruction favourable scores were significantly different between majors and non-majors in 2010-11 (p=0.016). In 2011-12 and 2012-13 there existed no statistically significant difference between major and non-major students in either the pre- or post-instruction results, with both cohorts showing almost no change over the academic year as shown in Figures 7.3 (b) and (c).

Table 7.3: Percentage of pre- and post-instruction favourable and unfavourable expert-like thinking for major and non-major students. The percentage favourable and unfavourable expert-like thinking do not necessarily add up to 100% due to the ability of students to choose a neutral response to each statement. Numbers in the brackets represent the standard error on the mean.

	Pre-Instruction %				Post-Instruction %			
Year	Majors		Non-Majors		Majors		Non-Majors	
	Fav	Unfav	Fav	Unfav	Fav	Unfav	Fav	Unfav
2010-11	73(2)	13(1)	69(2)	13 1)	72(2)	13(1)	64(2)	17(2)
2011-12	74(2)	10(1)	74(1)	10(1)	74(2)	11(1)	74(2)	11(1)
2012-13	72(2)	11(1)	70(2)	11(1)	70(2)	11(1)	70(2)	12(1)

Overall, with the one exception of 2010-11 post-test results, there were no attitudinal differences between majors and non-majors and therefore this does not help to explain the existence of the observed discrepancy between these two cohorts on the FCI. The results from the major and non-major students in this year do reflect the fact that when looking at the whole class cohort, 2010-11 showed a significant drop in expert-like thinking, unlike the 2011-12 and 2012-13 cohorts. It is also important, when comparing these results to other published studies, to remember that non-major students enrolled on the first year physics courses at the University of Edinburgh are very different from non-major students discussed in results from North American universities, who are not required to hold the same high school qualifications as physics majors. Results from the literature show that non-major students appear to have significantly less expert-like attitudes and beliefs towards physics compared to physics majors [178, 179].

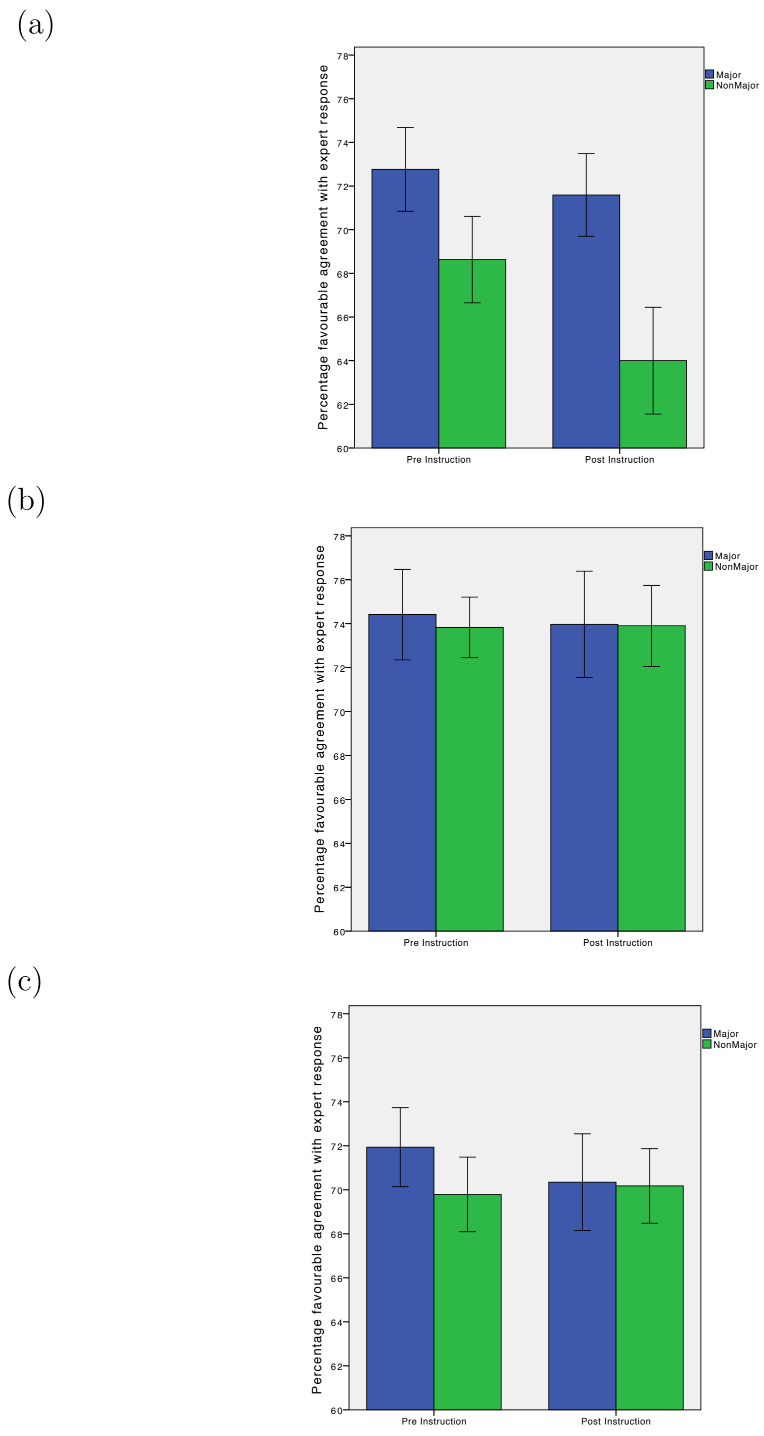


Figure 7.3: Percentage of favourable expert-like thinking pre- and post-instruction for (a) 2010-11, (b) 2011-12 and (c) 2012-13 major and non-major students. Error bars represent the standard error on the mean.

7.2.3 Comparison of male and female students' expert-like thinking

The matched data was split to investigate differences between male and female first year students. No differences were found between year groups for these two cohorts. In this section the pre-instruction and post-instruction expert-like thinking levels of males and females will be considered, looking at each academic year separately. Pre-instruction and post-instruction favourable and unfavourable percentage scores are shown in Table 7.4.

Table 7.4: Percentage of pre- and post-instruction favourable and unfavourable expert-like thinking scores for male and female students. Numbers in the brackets represent the standard error on the mean. For 2010-11 N(males)=71 and N(females)=33. For 2011-12 N(males)=54 and N(females)=22. For 2012-13 N(males)=43 and N(females)=19.

	Pre-Instruction %				Post-Instruction %			
Year	Males		Females		Males		Females	
	Fav	Unfav	Fav	Unfav	Fav	Unfav	Fav	Unfav
2010-11	70(2)	13(1)	72(2)	13 1)	66(2)	16(2)	72(2)	13(1)
2011-12	74(1)	10(1)	73(3)	11(1)	74(2)	11(1)	72(3)	11(1)
2012-13	71(1)	11(1)	70(3)	12(1)	70(1)	12(1)	70(4)	10(2)

In 2010-11 males entered the introductory physics course with a pre-instruction favourable percentage of 70(2)% compared to females who had a favourable percentage of 72(2)% ($p=0.689$), as shown in Figure 7.4 (a). At the end of the academic year males showed a 4% decrease to 66(2)% ($p=0.003$). In contrast, the expert-like response for females remained constant ($p=0.805$). For first year undergraduate students in the 2011-12 year group (Figure 7.4 (b)), there was no statistical difference between pre-teaching CLASS scores for males and females ($p=0.737$). Unlike in the previous year, male students showed no change in favourable score when presented with the survey at the end of the academic year. Female students' average favourable score decreased marginally from 73(3)% at the beginning of the year to 72(3)% at the end of semester two. Once again, there was no statistically significant gender gap post-test ($p=0.425$). Finally, considering 2012-13 students (Figure 7.4 (c)), males entered with 71(1)% agreement with experts and females entered with 70(3)% agreement ($p=0.815$). Female students showed no decline in expert-like thinking post teaching (p=0.941). Male students showed only a minimal decrease to 70(1)% (p=0.670).

By observing the percentage unfavourable scores for male students, it was seen that there was a statistically significant increase in disagreement with the expert responses for 2010-11 male students (p=0.015). Males in 2011-12 and 2012-13 showed no discernible change in average unfavourable scores. The unfavourable scores for female students remained constant in 2010-11 and 2011-12. Although the percentage disagreement decreased slightly for females in 2012-13, from 12(1)% to 10(2)%, it was not a significant drop (p=0.253).

These three years of data did follow the same trends seen in previous years [177]. As was found in the 2008-10 data, there existed no statistical differences between male and female students' attitudes as they entered the university degree programme (p=0.689 for 2010-11, p=0.737 for 2011-12 and p=0.989 for 2012-13). Where in previous years a significant gender difference existed post-instruction, there were no such differences between male and female students in any of the three academic years (p=0.063 for 2010-11, p=0.425 for 2011-12 and p=0.986 for 2012-13). In contrast to the measurable decline in expert-like thinking scores demonstrated in previous years, the three years presented in this chapter showed no significant decline in male and female attitudes after two semesters of teaching.

7.2.4 Comparison of male and female major and non-major students' expert-like thinking

The academic cohort was split by both gender and degree major in order to determine if degree intention has an observable effect on the physics attitudes and beliefs of male and female students. Table 7.5 indicates the number of students in each year group as a function of their gender and major. The proportion of major and non-major students in the matched data set was approximately equal in 2010-11 and 2011-12. In 2012-13 there was a higher proportion of non-major students, particularly in the male cohort, than physics majors. In 2011-12 the female population had a higher percentage of major students, although absolute numbers were very small.

Pre- and post-instruction favourable expert-like thinking scores for male and female major and non-major students in each academic year are shown in Table 7.6. Female majors in 2010-11 had a significantly higher level of

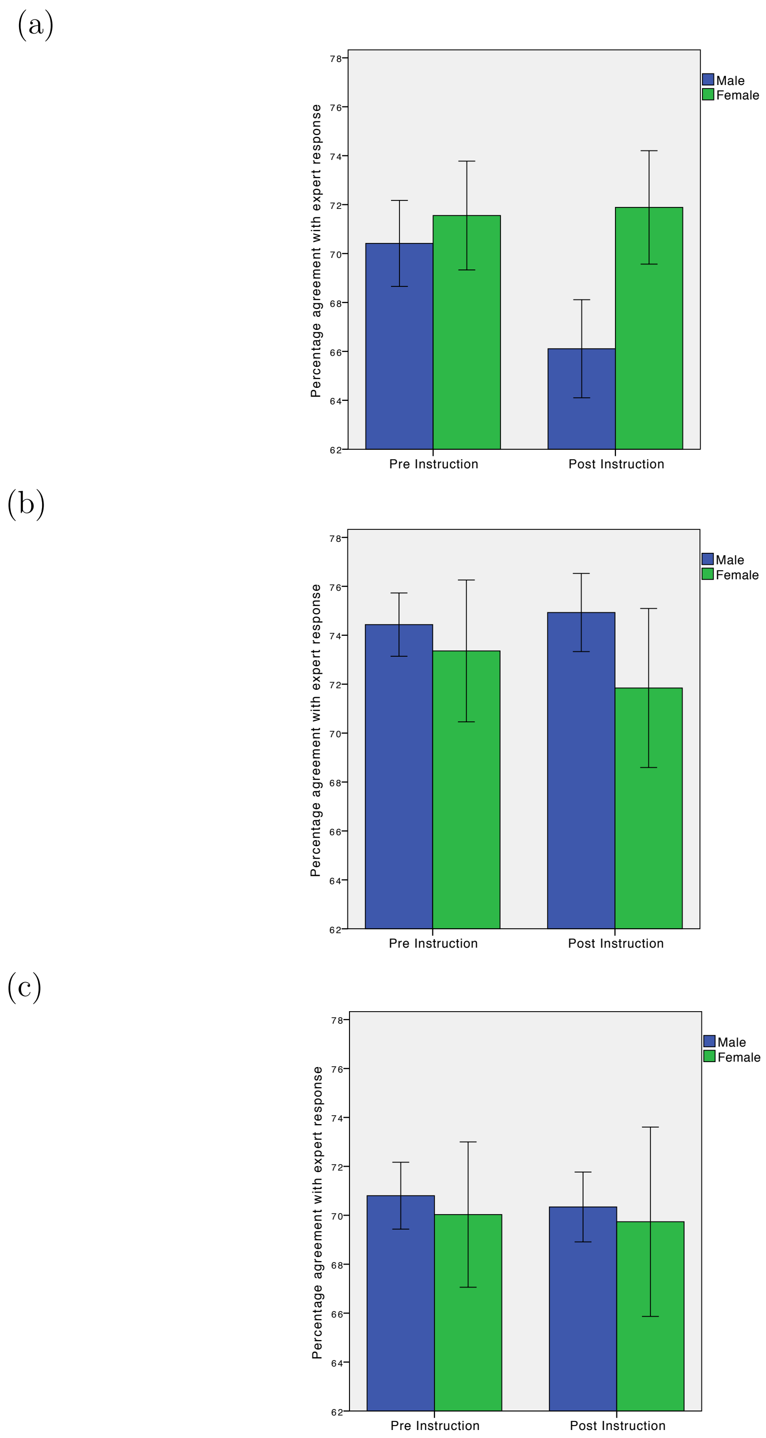


Figure 7.4: Percentage of favourable expert-like thinking pre- and post-instruction for (a) 2010-11, (b) 2011-12 and (c) 2012-13 male and female students. Error bars represent the standard error on the mean.

Table 7.5: Number of students as a function of subject major and gender who completed both pre-instruction and post-instruction CLASS surveys.

Year	Male Majors	Male Non-Majors	Female Majors	Female Non-Majors
2010-11	37	34	17	16
2011-12	25	29	13	9
2012-13	43	61	8	11

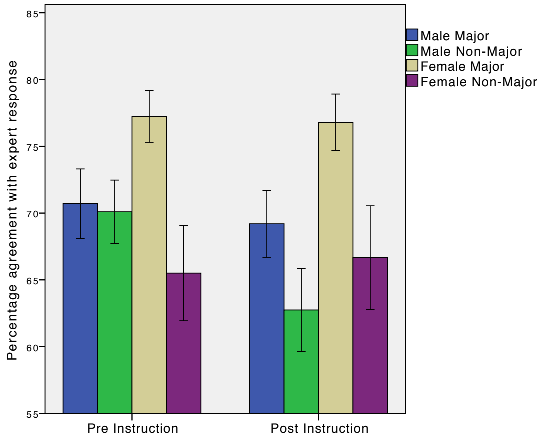
expert-like thinking, compared to all other cohorts as shown in Figure 7.5 (a) (compared to male majors p=0.049, compared to male non-majors p=0.024 and compared to female non-majors p=0.008). The gap between major and non-major females was particularly evident. Majors scored 77(2)% and non-majors scored 66(4)%. In the post-instruction results this gap remained significant (p=0.031). Interestingly, female majors showed no change in their attitudes whilst all male groups decreased. The male non-majors showed a significant decrease of 7% (p=0.003).

Table 7.6: Percentage of pre- and post-instruction favourable expert-like thinking scores for male and female major and non-major students. Numbers in the brackets represent the standard error on the mean.

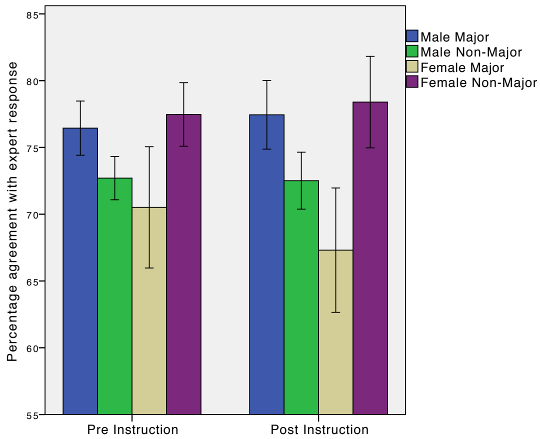
	Pre-Instruction %				Post-Instruction %			
Year	Majors		Non-majors		Majors		Non-majors	
	Males	Females	Males	Females	Males	Females	Males	Females
2010-11	71(3)	77(2)	70(2)	66(4)	69(3)	77(2)	63(3)	67(4)
2011-12	76(2)	71(5)	73(2)	78(2)	77(3)	67(5)	73(2)	78(3)
2012-13	72(2)	71(4)	70(2)	69(5)	71(2)	67(8)	70(2)	72(3)

In contrast, the 2011-12 academic year showed no significant differences between any cohorts in either the pre- or post-instruction results. In addition, each group showed very little change from the beginning to the end of the academic year (Figure 7.5 (b)). This was also the case in 2012-13, where there was no measurable difference between the populations (Figure 7.5 (c)). The size of the cohorts should be taken into account when drawing conclusions from this data. In particular, there are very small numbers of female majors and female non-majors in each of the individual year groups, resulting in relatively large

(a)



(b)



(c)

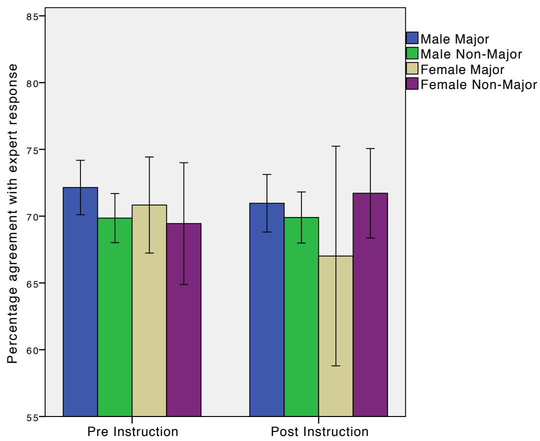


Figure 7.5: Percentage of favourable expert-like thinking pre- and post-instruction as a function of gender and major for (a) 2010-11, (b) 2011-12 and (c) 2012-13 first year undergraduate students.

standard errors on the mean.

7.2.5 Results of CLASS categories

As well as the overall expert-like thinking scores, which take into account all scored questions on the CLASS survey, analysis of scores from the eight question categories can provide a greater insight into students' attitudes and thinking. Conducting analysis at this level can help determine if the change in students' attitudes is related to a change in a specific area of their expert-like thinking which would not be highlighted in their overall CLASS scores previously presented. As mentioned previously, 27 of the survey questions have been placed into one or more of these eight categories listed in section 7.1. Pre- and post-instruction favourable category results for the whole class cohort in each academic year are shown in Table 7.7.

Table 7.7: Percentages of favourable pre- and post-instruction CLASS responses for the overall survey and the eight question categories. The numbers in the brackets represent the standard error on the mean.

Categories	2010-11		2011-12		2012-13	
	Pre	Post	Pre	Post	Pre	Post
Overall	71(1)	68(2)	74(1)	74(2)	71(1)	70(1)
Personal Interest	76(2)	73(3)	78(2)	80(3)	73(2)	73(2)
Real World Connection	**73(3)**	**75(3)**	**76(3)**	**80(3)**	**74(3)**	**77(1)**
PS General	78(2)	72(2)	84(2)	79(2)	77(2)	75(2)
PS Confidence	75(3)	71(3)	83(2)	76(3)	71(3)	70(2)
PS Sophistication	69(3)	63(3)	76(2)	74(3)	70(2)	67(2)
Sense making / Effort	82(2)	76(2)	83(2)	81(2)	82(2)	76(2)
Conceptual Understanding	71(2)	65(2)	74(2)	76(2)	73(2)	72(2)
Applied Conceptual Understanding	59(3)	54(3)	63(2)	65(2)	61(2)	59(2)

Although little change was observed in the overall favourable scores between pre- and post-instruction, with the exception of the drop in 2010-11, there was a wide range of scores on a category level. Student scores between the beginning of the undergraduate degree and the end of their first academic year decreased in some categories but increased in others. A comparison of results between the three academic years highlighted a few interesting features. First, specific categories showed much higher or lower expert-like thinking scores than others. In particular, students' level of *'Sense Making/Effort'* was very high, with students beginning their degree with a favourable score of 82(2)% in 2010-11 and 2012-13 and 83(2)% in 2011-12. Conversely, their *'Applied Conceptual Understanding'* scores were relatively low, particularly in 2010-11. This category contains seven questions measuring students' view of their ability to transfer physics knowledge from one question to another. The ability to apply physics principles to questions set in different contexts is an important skill in physics problem solving.

In the majority of cases there existed a decrease in scores between the pre- and post-instruction survey results. In the case of the 2010-11 cohort, seven of the eight categories showed a negative change in expert-like attitudes. The exception to this was in the *'Real World Connection'* category. In fact, in all three year groups students' *'Real World Connection'* scores increased after two semesters of teaching. This increase was not, however, statistically significant. A study by Milner-Bolotin at a Canadian university saw a positive shift in expert-like thinking in all eight categories including the *'Real World Connection'* category [180].

There were also some gender differences in the category scores (Table 7.8). In particular, males outperformed females in the *'Problem Solving Sophistication'* category in both 2011-12 and 2012-13 at the beginning of semester 1. In 2011-12 males had a pre-test favourable score of 79(2)% compared to females who had a favourable score of 70(6)%, but this difference was not significant ($p=0.162$). Similarly, in 2012-13 males had a pre-test score of 72(2)% compared to females who scored 60(6)% agreement with the expert responses ($p=0.078$). Looking at the *'Real World Connection'* category, male and female students had equal levels of expert-like thinking in 2010-11 and 2012-13. In 2011-12 female students had a much higher percentage favourable score compared to male students, although this was not significant at the 95% level ($p=0.154$). A study using an Arabic version of the CLASS survey in Saudi Arabia amongst introductory physics students also looked at gender differences at a category level [85]. They noted that women, particularly those on pre-med courses, tended to have equal or higher scores in the *'Personal Interest'* category and demonstrated more expert-like beliefs about physics than men overall. This is in slight contrast to our results in which male and female students showed no difference in their levels of *'Personal Interest'*.

Table 7.8: Percentages of favourable pre-test CLASS responses for the overall survey and the eight question categories as a function of gender. The numbers in the brackets represent the standard error on the mean.

Categories	2010-11		2011-12		2012-13	
	Males	Females	Males	Females	Males	Females
Overall	70(2)	72(2)	74(2)	73(3)	71(1)	70(3)
Personal Interest	76(3)	75(4)	78(3)	77(5)	73(2)	75(5)
Real World Connection	73(3)	73(4)	73(4)	83(6)	74(3)	74(7)
PS General	78(3)	79(3)	85(2)	82(4)	77(2)	78(5)
PS Confidence	74(3)	77(2)	84(3)	81(5)	70(3)	71(7)
PS Sophistication	**68(3)**	**70(5)**	**79(2)**	**70(6)**	**72(2)**	**60(6)**
Sense Making / Effort	83(2)	82(3)	84(2)	78(5)	82(2)	81(3)
Conceptual Understanding	70(3)	74(3)	75(3)	74(5)	73(2)	70(5)
Applied Conceptual Understanding	57(3)	62(4)	63(2)	62(5)	62(2)	54(5)

Each category defined in the CLASS survey contains very few questions, with the maximum number of questions included in any category being eight and the minimum being four. This means that a shift in favourable or unfavourable responses to an individual statement can result in a large change in the overall category score. This adds a degree of caution with which the comparison of category scores between two different cohorts should be treated. In particular, the *'Real World Connection'* category comprises only four statements.

7.2.6 Summary of CLASS results

Results have shown that students entering the first year of the undergraduate physics programme at the University of Edinburgh have relatively high levels of expert-like thinking. Although the 2010-11 year group showed a statistical decrease in this level of expert-like thinking, this was not the case for the following two academic years, perhaps suggesting that teaching methodologies implemented in the first year courses have had a positive effect on maintaining students' expert-like thinking. No significant discrepancies were found between the attitudes and beliefs of major and non-major physics students at the beginning of the semester. In two of the three years examined this remained the case after two semesters of teaching. Similarly, there were no differences between male and female cohorts in either the pre- or post-instruction survey results. This was in direct contrast with results seen prior to 2010, in which females had significantly lower levels of expert-like thinking than males at the end of the academic year. Analysis indicated that students' levels of agreement with the US 'expert' responses differed depending on the question category. In particular, students had very low levels of *'Applied Conceptual Understanding'*.

7.3 UK academics' attitudes towards learning and studying physics

One of the key features of the CLASS survey, in comparison with other attitudinal tests, is its ability to not only provide a quantitative measure of the change in students' attitudes over a defined period of time, but to compare responses from students with those of 'experts'. As discussed in Chapter 2, during the design of the instrument the survey statements were presented to physics faculty members at the University of Colorado. A total of 16 academic staff completed the survey and their responses were used to create the 'expert' opinion against which students are scored. In situations where no consensus was reached between staff responses, the statement was further discussed amongst the staff to try and agree on a favourable or unfavourable response. Six statements were not given an expert response, but remain in the survey as unscored statements. This section details results from work carried out to compare attitudes of male and female academics, industry members and people at different levels of academia in the UK. Comparisons were made between responses from academics in the UK and the US academics who participated in the original survey validation. To the best of our knowledge, this is the first time this instrument has been used to test the attitudes of a range of physics academics across the UK.

7.3.1 Methodology and demographics of respondents

The CLASS survey was distributed to members of the Institute of Physics (IOP) in order to gain a measure of expert views from physics graduates across the UK. The survey was distributed in May 2011 using the online survey tool 'Survey Monkey'. In addition to answering the survey, participants supplied demographic information including their gender, current role of employment (level of academic role or whether they were currently working in industry) and the number of years since the completion of their first degree. Each participant answered the survey only once rather than twice as in the pre- and post-test methodology. This means that results are representative of the current level of expert-like thinking of each sub-group, rather than a measure of their change in agreement of expert-like thinking over a period of time.

A total of 421 completed surveys was collected, with respondents coming from a diverse range of backgrounds: academics[1], postdoctorates (PDRAs), postgraduates and industry members (as shown in Table 7.9). Of these participants, 75% were male and 25% were female. The numbers in each group do not add up to the total number of respondents as some people did not provide information concerning their current role of employment or academic position. The percentage of male academics (lecturers, senior lectures, professors) was much higher than that for postdoctorates or postgraduate researchers. Only 16% of academics who completed the survey were female, reflecting the national average for all physics academic staff in higher education institutions in the UK (17%) [181]. When looking at male and female academics, it was seen that 69% of males and 58% of females had 20 years or more experience since the completion of their first degree. A further 29% of males and 38% of females indicated it was 11-16 years since the completion of their degree.

7.3.2 CLASS responses as a function of academic background

Table 7.10 shows the percentage favourable expert-like thinking scores for each employment level. Moving up the academic career level, results showed an overall

[1]The term academics used in this study refers to those working in academia at a level higher than postdoctorate (i.e. lecturers, senior lecturers, professors).

Table 7.9: Number of survey responses from IOP members as a function of gender and employment level.

	N	N(Males)	% Males	N(Females)	% Females
All	421	315	75%	106	25%
Academics	160	135	84%	25	16%
PDRAs	56	36	64%	20	36%
Postgraduates	115	78	68%	37	32%
Industry Members	53	41	77%	12	23%

trend of increased CLASS scores. In particular, academics had a higher favourable score compared to all other groups, with a favourable score of 86(1)%. Comparing academics to PDRAs indicated a statistically significant difference (p=0.014), as was the difference between academics and postgraduates (p<0.001). Members of the IOP not currently working in academia had an expert-like thinking score of 81(1)%. Once again, this was found to be significantly lower than that of academics (p=0.004).

Table 7.10: Percentage of favourable and unfavourable expert-like thinking scores for each employment level group. Numbers in the brackets refer to the standard error on the mean.

	N	Favourable %	Unfavourable %
All	421	82(1)	7(0)
Academics	160	86(1)	5(0)
PDRAs	56	82(1)	7(1)
Postgraduates	115	80(1)	9(1)
Industry Members	53	81(1)	7(1)

It is not necessarily the case that we should expect all experts to have a single and consistent view for all items on the survey. Therefore it is not necessarily expected that respondents will score 100% agreement with the pre-defined expert response. However, looking at the percentage disagreement with the expert responses, it was seen that all populations had a lower percentage unfavourable score compared to first year undergraduate students at Edinburgh, as seen in Figure 7.2.

7.3.3 CLASS responses as a function of gender

Of particular interest was whether there existed any gender differences in responses within each employment group. Data was subsequently split by gender and average CLASS scores calculated for each cohort. Results of favourable agreement scores are shown in Figure 7.6.

When analysing all 421 responses to the survey as a function of gender, there existed no statistical difference between male and female overall favourable scores, with males scoring 82(1)% and females scoring 81(1)%, nor were there any significant gender differences on individual categories. The one exception to this was for the category of *'Applied Conceptual Understanding'* which contains seven questions probing how students solve problems (for example their use of formulae or application of similar strategies to solve different problems) and whether they feel that physics consists of *'many disconnected topics'*. Here males showed a significantly higher level of expert-like thinking than females (p=0.003). Male respondents had an average favourable score of 75(1)%, compared to 67(2)% for females.

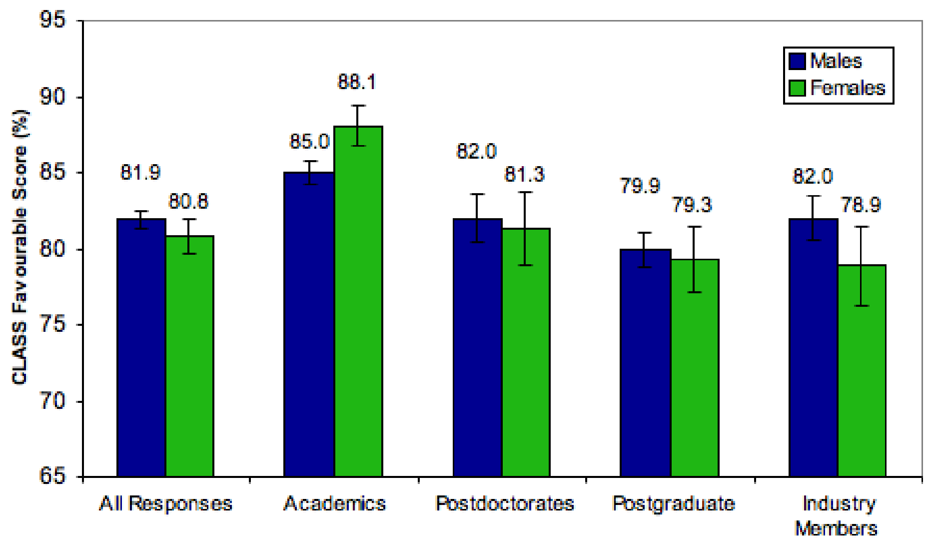


Figure 7.6: Percentage favourable expert-like thinking scores as a function of gender and employment level. Error bars represent the standard error on the mean. Numbers above each bar represent the percentage agreement of expert-like thinking for each population.

There were also no significant gender differences for PDRAs (p=0.312), postgraduates (p=0.820) or industry members (p=0.302). There was, however, a significant difference between male and female academics (p=0.047), with female academics having a significantly higher favourable expert score than male academics. The distribution of these scores is shown in Figure 7.7. Male academics had a much larger range of favourable scores than female academics. Male academics had an overall favourable score of 85(1)%, compared to female academics who had an overall score of 88(1)%. This was the only cohort in which females were more expert-like in their responses than males

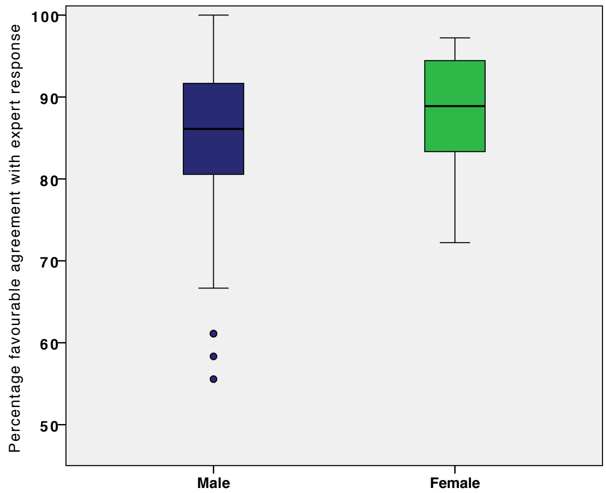


Figure 7.7: Boxplot of male and female academics' overall percentage favourable scores.

What is particularly intriguing was that female academics scored more highly than male academics in all eight of the survey categories. Percentage favourable scores for each category can be seen in Figure 7.8. For the *'Real World Connection'* category this difference was statistically significant (p=0.026). Female academics showed significantly higher agreement (94.2%) with the expert response than male academics (88.6%). This category contains four statements probing to what extent participants use personal experiences or real world situations to further their understanding of physics. This result supports the hypothesis made earlier about female undergraduates relating physics principles to 'real world' situations to a greater extent than males. When we consider only respondents who identified themselves as working in industry, we found that there were no differences between male and female attitudes across all eight categories.

This was also the case for male and female postdoctorates and postgraduates.

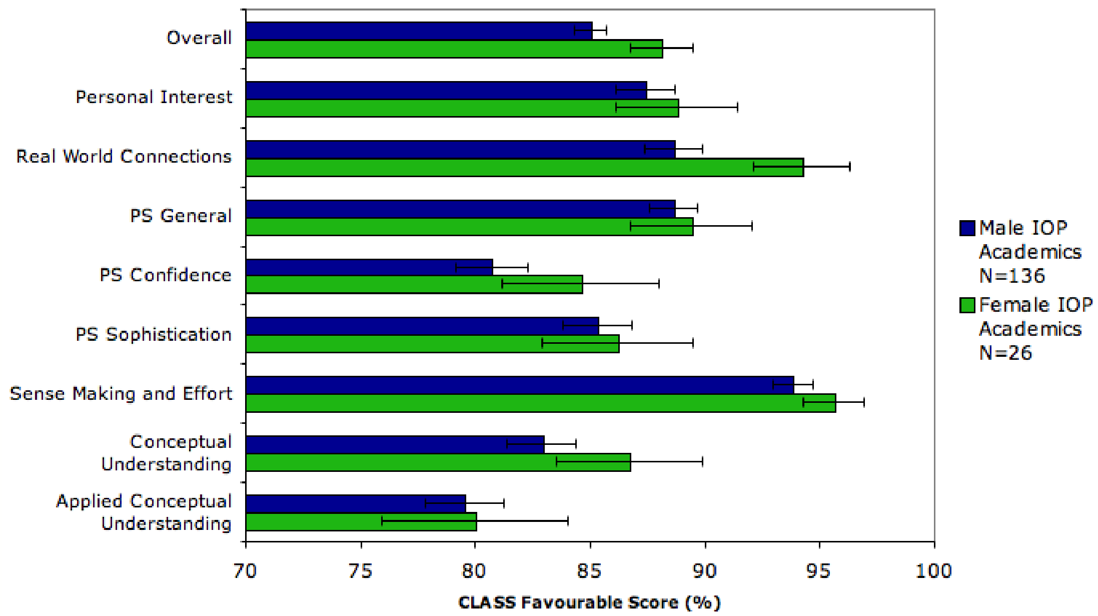


Figure 7.8: Male and female UK academics' favourable percentage scores on each of the eight CLASS categories. Error bars represent the standard error on the mean.

7.3.4 Questions showing gender differences between academics

The results from undergraduate students, discussed in section 7.2, compared the expert-like thinking of first year undergraduates with those of a physics faculty at an American university. Having identified a significant difference between the level of expert-like thinking of male and female UK academics, it was queried whether this overall discrepancy between cohorts would have the effect of altering the expert agreement or disagreement response to any of the individual CLASS statements. If the expert response to a particular statement was different between these cohorts this could have the effect of altering the 'Agree' or 'Disagree' response against which students are graded, thereby having an effect on their overall score.

The percentage of IOP members who identified themselves as an academic, who agreed, disagreed or remained neutral for each statement was calculated to determine if there was a lack of consensus (or even disagreement) with the original expert view. A level of consensus of 2/3 (67% agreement with the expert view) was chosen as the threshold to identify questions for which it was deemed there was not a broadly consistent view amongst our survey population of academic staff. Applying this criterion, it was found that the responses by UK and US academic staff to items on the CLASS survey were not consistent for five of the statements, each of which will be discussed below.

CLASS statement 5

One example of a question falling below this threshold was statement 5 of the survey:

"After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic."

The 'expert' response to this statement, as defined by US faculty members is 'Disagree'. Responses from UK academics indicated that only 65.2% of participants disagreed with the statement (i.e. agreed with the 'expert' response). A further 14.3% agreed with the statement and 20.5% chose to remain neutral. When we look at these responses as a function of gender, a slightly higher percentage of female academics agreed with the US 'expert' response (68.0%) compared to male academics (64.7%). This question belongs to three statement categories: *'Problem Solving Sophistication'*, *'Conceptual Understanding'* and *'Applied Conceptual Understanding'*.

In order to answer the question of whether undergraduates at the University of Edinburgh look more like UK experts than US experts, the percentage of male and female first year 2012-13 students who agreed or disagreed with each of these statements at the beginning of semester one was determined. For statement 5, 65.4% of male undergraduates and 42.1% of female undergraduates disagreed with the statement (i.e. agreed with the 'expert' response). The distribution of male responses across the 'Agree', 'Neutral' and 'Disagree' statements was very similar to that of the UK academics. Female students were more evenly distributed across all possible answers, with 26.3% agreeing with the statement and 31.6% remaining neutral.

CLASS statement 12

Statement 12 of the attitudinal survey states:

"I cannot learn physics if the teacher does not explain things well in class."

The 'expert' response defined for this statement is to 'Disagree'. Only 60.9% of UK academics disagreed with this statement. Once again, the percentage of respondents who chose a neutral agreement to this statement was relatively high (21.7%), with the remaining 28 participants (17.4%) choosing to disagree with the predefined expert response. There was a large discrepancy between the opinions of male and female UK academics. Only 59.6% of males had the same opinion as US faculty members, compared to 68.0% of females. This higher percentage of female agreement with the expert response may suggest that female instructors are more inclined to believe that students will complete reasonable amounts of self-study and seek help if they do not feel they completely comprehend a subject. Question 12 did not belong to any CLASS category.

Undergraduate students' responses to this statement were very different from those of the UK and US experts. Only 21.2% of males chose to 'Disagree', with 38.5% answering 'Agree'. Female undergraduates in 2012-13 were even more unlikely to agree with the expert response. Only two of the nineteen female students (11%) chose to 'Disagree' with the statement and 63.2% chose to 'Agree' that they couldn't learn physics unless it was well explained by the instructor.

CLASS statement 14

Participants were asked to give their opinion to statement 14 which states:

"I study physics to learn knowledge that will be useful in my life outside of school."

Only 58.8% of academics agreed with the expert response which was 'Agree'. In this instance, both male and female academics from the IOP had a percentage expert-like response below the chosen threshold of 67%. In total, 57.8% of male academics and 64.0% of female academics agreed. This statement was an example of a statement from the *'Personal Interest'* category in which respondents are prompted to see if they have a personal connection to physics and its uses in everyday life. Although both genders showed relatively low agreement with the

US expert response, female academics did show a higher percentage agreement compared to male, suggesting that females may be more likely to try to relate physics to everyday situations, as was discussed in earlier chapters.

Once again, the responses of first year undergraduates in 2012-13 were compared to those of the UK and US academics. Both cohorts showed good levels of agreement with the UK experts, with 52.9% and 63.2% of male and female students answering 'Agree' to statement 14 respectively.

CLASS statement 16

Statement 16 of the survey belongs to two categories: *'Problem Solving General'* and *'Problem Solving Confidence'*. Respondents were asked to agree or disagree with the statement:

"Nearly everyone is capable of understanding physics if they work at it."

Interestingly, only 57% of physics academics agreed with the predefined 'expert' response which was 'Agree'. A further 24.4% of respondents chose to give a neutral reply, the remaining 18.6% disagreeing. The level of agreement with the expert response was particularly low for male academics in the IOP, only 54.1% of whom agreed with the statement. In contrast, 69.2% of females agreed with the predefined expert answer. This large gender discrepancy is perhaps surprising. One possible explanation may be that female academics are more aware of the need to invest in self-study if they are unsure of a certain concept and are therefore more diligent than males. From the perspective of an instructor, this could suggest that female academics are more inclusive of everyone's ability to learn physics, whilst some male academics may be less aware of the potential need for additional support for some students.

As with the UK academics, a higher proportion of female first year undergraduates than males chose to answer 'Agree' to the above statement. In total, 68.4% of females and 59.6% of males agreed with the expert response, suggesting that the undergraduates had similar attitudes to the UK population of academic staff.

CLASS statement 37

The final statement which did not show a consensus at the 67% level was statement 37 which states:

"To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed."

In total 66.3% of academics agreed with the US expert response which was 'Agree'. This survey item showed a very large gender discrepancy between the responses of UK academics. While only 62.7% of males shared the same response as the US physics faculty, 84.6% of females agreed with the 'expert' response. This question belonged to the *'Real World Connection'* category, and once again reiterates the idea of a female bias towards 'real world' comparisons.

The percentage agreement with the 'expert' response was much lower for undergraduate students than the UK academics. Only 51.9% of males and 57.9% of females chose to 'Agree' with statement 37. The higher percentage agreement of females relative to males was nevertheless reflected in the undergraduate responses.

Results from the above analysis revealed that the five statements which showed a lack of consensus by UK academics did not belong to a single statement category, but in fact were distributed amongst seven of the eight categories. There is no available information pertaining to the percentage of the original US faculty members who initially agreed or disagreed with each statement with which to compare UK data. During the survey validation process each expert response was determined through a straight consensus. It is therefore unclear which statements required further discussion amongst 'experts' for a consensus to be reached.

As part of the study, participants were encouraged to provide comments or feedback on the test questions after they had completed the survey. These comments highlighted several potential factors affecting participants' responses. One male academic commented that statement 22, which states:

"If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations."

does not specify what is meant by the term 'situations'. He commented that this statement could be ambiguous in its interpretation and said that *"If two different*

completely physical systems follow the same mathematical law, is that the same situation?".
Another question considered ambiguous by several respondents was statement 41:

> *"It is possible for physicists to carefully perform the same experiment and get two very different results that are both correct."*

One respondent commented that it would depend on whether statistical variation had been taken into account and:

> *"it is possible for physicists to carefully perform the same experiment and get two very different results that are both correct. I replied 'Agree', as it depends if the experiment has been successfully constructed to observe the variable of interest, controlling all others. If the experiment is not controlling all factors, then there could be natural unexplained variation."*

One academic commented further saying that theorists and experimentalists may have differing expert responses to some of the statements. This level of detail was not recorded during this study but offers a potential area of investigation for future research. It should be recognised that, although this survey was deliberately administered to academics and industry members who have completed their undergraduate education, several of the survey statements are targeted towards students and their experiences of school and undergraduate teaching, as can be seen by the language used in the statements.

7.3.5 Summary

There is evidence to suggest that there exist some gender differences for academics on both category and individual item responses, suggesting that women in academia may have different attitudes and beliefs from men about their subject. These results also indicated that the expert view for some items of the CLASS survey, as measured by responses from UK academics, is not the same as that of the US faculty members used to validate the original survey. In fact, for academics, females had a higher expert-like thinking score than males. This research has been undertaken using a much larger cohort of faculty than when the instrument was originally validated. Whilst there existed statements for which there appear to be a lack of consensus on the defined 'expert' view, these did not have the effect of changing the defined expert response for any statement from 'Agree' to 'Disagree' or vice versa. For this to happen the level of agreement with the predefined expert response would need to be below 50% on an individual item. Nevertheless the differences in academic attitudes highlighted in this study may have future implications for how we should calibrate student responses to such a survey. The fact that UK academics did not score 100% agreement with the view of US physics faculty suggests that it may not necessarily be sensible to expect our students to achieve this level of expert-like thinking.

The methodology of this process was somewhat different to that undertaken by US faculty members, as UK academics were not given the opportunity to discuss their opinions with other academics in order to reach a consenus. Respondents' comments suggested that some of the survey questions are ambiguous in their meaning and could cause discrepancies in the way in which they are interpreted. It was also commented that some of the questions relating to students' opinions of how physics is presented in a classroom are not relevant to those no longer in high school or university. This may also have had an affect on how academics responded to such questions. It needs to be considered whether the variations witnessed in academics' responses are the result of variations in academic attitudes or whether they are the result of structure of the survey instrument.

7.4 Undergraduate physics students' intentions

In order to gain a better understanding of why students choose to undertake a physics degree and whether this reasoning differs for male and female students, first year undergraduate students enrolled on the 2010-11 introductory physics course ('Physics 1A') were presented with a paper survey asking them to specify their current degree programme, their highest secondary school physics qualification and their intended degree exit point (BSc, MSc[2] or PhD). Surveyed students were also asked to comment on up to three reasons for pursuing their chosen degree. The survey was undertaken during the first week of formal teaching, so it may be necessary to consider the fact that university was

[2]The first year 'Physics 1A' population comprises students from both physics and other degree programmes and therefore the MSc qualification can refer to both the Integrated Masters in the School of Physics and Astronomy or to a Masters qualification from another discipline.

a new experience and their only exposure to the subject had been through their experiences of school teaching methods. Students in other years of the undergraduate programme were also surveyed and asked about their school physics qualifications and intended degree exit point, creating a cross-sectional view of students opinions throughout the undergraduate degree.

7.4.1 Intended degree

Students completing the survey were asked to indicate the degree programme on which they were currently enrolled from a choice of six disciplines: Physics, Mathematics, Chemistry, Informatics, Engineering and Other. The flexibility of the degree system at the University of Edinburgh enables students to change their degree course during or after their first year of study, depending on the courses taken during this first undergraduate year. Consequently, the results of this survey indicated the distribution of the 'Physics 1A' cohort across different degree programmes at the start of their university career and may not be representative of the distribution after one year of study.

Of the students who replied, 50% were intending to study towards a Physics degree, as shown in Figure 7.9. Chemistry and Engineering were also well represented, with 14% and 15% of students respectively. The introductory course comprised 10% Informatics students and 7% Mathematics students, with the remaining respondents enrolled on other degree courses. When looking in detail at the gender differences in degree choice, it was seen that Engineering, Mathematics and Informatics students taking 'Physics 1A' were more likely to be male. A higher proportion of the female cohort was enrolled on a Physics or Chemistry degree compared to males. This is in contrast to the overall gender participation gap in favour of males on the physics degree programme and is most likely a result of the fact that, of the 207 students enrolled in 2010-11, only 115 answered this question. This also meant that the absolute number of students in each of these categories was relatively small.

In addition to their degree choice students were asked *"How sure are you that you will graduate in that discipline?"*. Answers were coded on a five point Likert scale (Very Sure, Sure, Neutral, Unsure and Very Unsure). Results showed that the vast majority of students had a positive conviction that they had chosen the correct degree discipline. Of those who completed the survey, 69% said that they

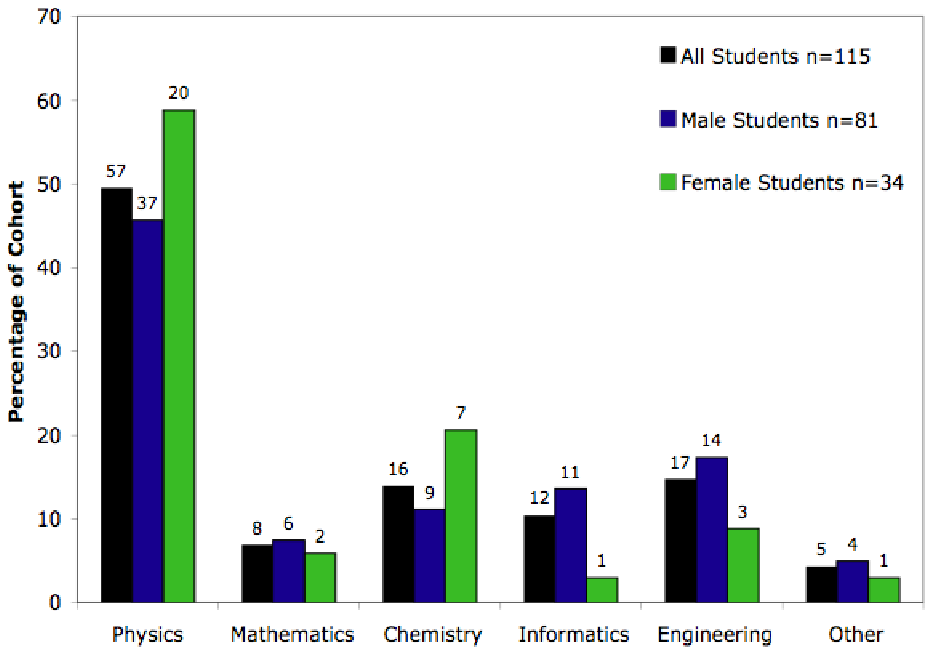


Figure 7.9: Percentage of male and female first year students enrolled on degree programmes in each discipline. N=115, N(males)=81 and N(females)=34. Numbers above each bar represent the absolute number of students in each group.

were either 'Very Sure' or 'Sure' that they would graduate in that discipline. Only 9% of students were either 'Very Unsure' or 'Unsure', with the remaining 22% choosing to remain 'Neutral'. While male students who answered favourably were split almost equally between 'Very Sure' and 'Sure', female students were much more likely to choose 'Sure'. Only one male student indicated that they were 'Very Unsure'.

7.4.2 Why did students choose their intended degree?

Students were presented with a blank survey and asked to write down up to three reasons why they chose to study their degree course. The total number of completed survey responses was 154 (N(male)=108 and N(female)=46). The total number of comments written by students was 364, some students providing three reasons and others only one or two comments. The comments were subsequently grouped into nineteen categories, each defining a different reason for choosing to study their degree. Any flippant or irrelevant comments were excluded from the analysis. The categories and corresponding number of comments, ranked in order of popularity, are shown in Table 7.11.

Table 7.11: Reasons for choosing to study intended degree ranked in order of popularity. N(male)=108 and N(female)=46.

Category	Number of Comments		
	All	Males	Females
Enjoyed subject at school	73	49	24
Interest	61	43	18
Job or financial prospects	50	34	16
Flexibility of degree/ Combines well with other subject	29	20	9
Good at it / Find it easy	28	22	6
Greater understanding of how things work	20	13	7
Research interests	17	14	3
Subject is applicable/ Relevant	15	12	3
Like maths	15	11	4
Challenging / Problem Solving	10	8	2
Travel opportunities	8	6	2
Course content	8	6	2
Unsure	7	7	0
Experiments	6	4	2
Career advice / Work experience	5	5	0
Interest in computing/ Programming	5	4	1
Reputation of degree	3	2	1
Reputation of department	2	2	0
Logical thinking	2	0	2

By providing them with blank pieces of paper and offering no guideline as to potential comments or categories, students were able to freely express their views. The most popular category, making up 20% of all comments, referred to students' enjoyment of the subject whilst at secondary school. Stating that they found it interesting and that the subject would offer promising job prospects upon completion of their degree were also very popular statements amongst students as a whole, written by 61 (17%) and 50 (14%) students respectively.

Comparisons were made between male and female reasons for choosing their degree programme to determine whether any gender differences existed. If gender differences were to exist this may go someway to help explain the existence of the gender participation gap. However, the three most popular categories were the same for males and females. Comments about their enjoyment of studying the subject at school made up 19% of male comments and 24% of female comments. Their interest in the subject was reflected in 16% of male comments and 18% of female comments. Both cohorts listed future job and financial prospects as a top reason for their choice, although males took this into consideration more (30% for males and 16% for females). Interest in current scientific research was a relatively popular comment, particularly amongst male students. This may reflect the fact that there was a higher proportion of males who stated they intended to pursue a PhD qualification in section 7.4.4.

The comments reflected the students' choice in their degree discipline, rather than in choosing to take 'Physics 1A' as a course subject. Therefore categories of response may be linked to the specific degree programmes. For example, almost all the eight students who referred to *'Travel opportunities'* were geophysicists, with two engineering students also stating that they saw their degree as an opportunity to pursue a *"job/career that may involve a lot of travelling"*. Many non-majors, in particular those pursuing an engineering degree, commented on the fact that physics complimented their degree choice and that they enrolled on 'Physics 1A' because of the flexibility allowed by the university curriculum to choose an outside course. Encouragingly, many of the students who discussed current scientific research, or the reputation of the degree or department, identified themselves as physics students. They commented on specific developments in research, such as experiments taking place at CERN, as well as their personal desire to enter the research community.

7.4.3 What are students' highest physics qualifications?

Research has suggested that students' background and grounding in physics and mathematics at school level may influence their commitment to physics [182]. Students in each year of the undergraduate degree programme were asked to state their highest physics qualification achieved prior to university: Higher, Advanced Higher, A-level, International Baccalaureate, Irish Leaving Certificate or Other. As discussed in Chapter 3, Scottish students usually have a background of Higher or Advanced Higher physics, whilst English students in general leave school with A-level physics. The first year course also has approximately 25% international students. At Edinburgh, approximately 40% of students who begin their physics degree withdraw or transfer to another degree course before completion of the BSc programme [97]. One hypothesis is that, as they move up through the years in the undergraduate programme, the students who withdraw are more likely to be those who entered with only Higher physics, whilst the ones who remain until 5th year are those who completed A-level or Advanced Higher physics qualifications. This hypothesis was explored by surveying students in each year of the degree programme to gain a cross-sectional view of students' highest physics qualifications. Results from this study are shown in Figure 7.10.

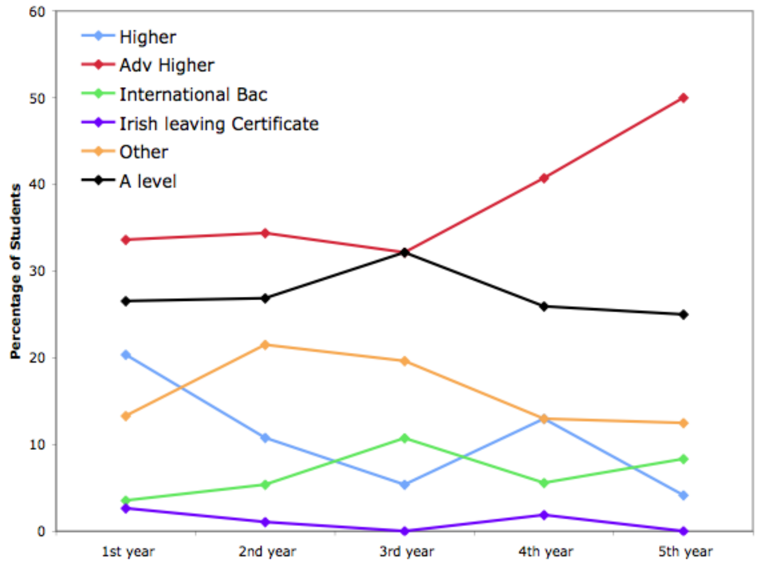


Figure 7.10: Highest physics qualifications of undergraduate students enrolled in each year of the undergraduate programme in 2010-11. N(first year)=113, N(second year)=93, N(third year)=38, N(fourth year)=47 and N(fifth year)=24.

When looking at the range of qualifications first year students have upon entering university, Advanced Higher physics was the most highly represented school qualification, completed by a third of students. A-level students were also very well represented, with 27% of students, followed by Higher physics with 20%. Over 13% of first year students chose 'Other' as their latest physics qualification, suggesting that they belong to the international student cohort. If results from only students whose intended degree is physics are considered these relative proportions remain very similar. In total 36% of these respondents completed Advanced Higher, 28% completed A-level, with a slightly lower percentage of Higher students (13%).

Results from the same survey presented to second year students showed a similar pattern. Once again, Advanced Higher physics was the most popular physics qualification amongst students (34%) followed by A-level physics (27%). The distribution of students amongst the possible school physics qualifications was very different for 3rd year undergraduates. After surveying students in the junior honours year group it was found that an equal number of students (32%) had come from a background of Advanced Higher and A-level physics. Only three of the thirty-nine third year students surveyed came from a Higher physics background. Similarly, in senior honours only seven students (13%) entered the physics degree programme with Higher physics. The proportion of students coming in with Advanced Higher physics had increased to 41%, whilst A-level physics continued to be the second most popular qualification. In the fifth year of undergraduate only 24 students replied to the survey. Only one of these students left school with Higher physics. Half of the students left with Advanced Higher physics, with a further quarter leaving with an A-level physics qualification. International Baccalaureate and students with 'Other' qualifications were still represented by a handful of students.

There existed some gender differences in each year group. In first year, Advanced Higher physics was the most represented qualification for males (36%) followed by A-level physics (25%). This differed from the female cohort who had an almost equal number of respondents who entered with Advanced Higher and A-level qualifications. This trend was reflected in responses from second year students. Once again, an almost equal number of female students completed Advanced Higher and A-level qualifications. Advanced Higher physics remained the most popular response for male students. Only fifteen third year students were female. Of these, seven completed Advanced Higher physics,and four completed A-level physics, with the remaining students split between the International Baccalaureate and Other. Of the forty-one male students in that year group, fourteen studied A-level physics before university, with a further eleven completing Advanced Higher physics. The small sample size of female cohort, and the obvious reduction in absolute numbers of females seen after the completion of 4th year, means that very little can be inferred about the trends regarding females.

The implications that students entering with Higher Physics are the least likely group to progress beyond BSc suggest that there may be the need to recognise this in first year, where there may exist a gap of knowledge compared with those entering with Advanced Higher or A-level physics. The percentage of students whose latest physics qualification was Higher Physics showed a significant decline as we move up the undergraduate year groups. A fifth of students entered university with a Higher qualification, but this decreased dramatically to only 10.8% by second year and decreased further in junior and senior honours. The percentage of Advanced Higher physics students remained fairly consistent during the first three years of undergraduate study, increasing by fourth and fifth year where it represented 50% of all MPhys students. The representation of A-level students remained consistently high throughout undergraduate years (25-26% of students).

7.4.4 Students' intended exit point

Students in each year group were asked whether they were intending to leave university after a Bachelors, Masters or PhD degree. By gaining a cross-sectional picture of the percentage of students considering each exit point as a function of undergraduate year of study (Figure 7.11), we are able to develop a better understanding of the changes in the motivation of students towards studying university level physics.

When surveyed in their first week of teaching, 46% of first year students indicated an intention to complete a Masters degree, with a further 21% intending to complete a BSc qualification. One limitation of the survey was that it did not differentiate between the intention to complete the Integrated Masters (MPhys) in physics and a postgraduate Masters (MSc). The percentage of women intending to complete the Masters programme (51%) was higher than for males (43%). This was also the case in the second year undergraduate results. Once again, the proportion of females intending to do a Masters (63%) was much higher than that of the male population (51%). The BSc was the most unpopular option for the whole class population, chosen by approximately 20% of students in both years.

When physics students began their undergraduate education a surprisingly high proportion of students (33%) considered remaining in research after their degree to pursue a PhD. The percentage of males intending studying for a PhD (38%) was quite a bit higher than that for the female population surveyed (23%).

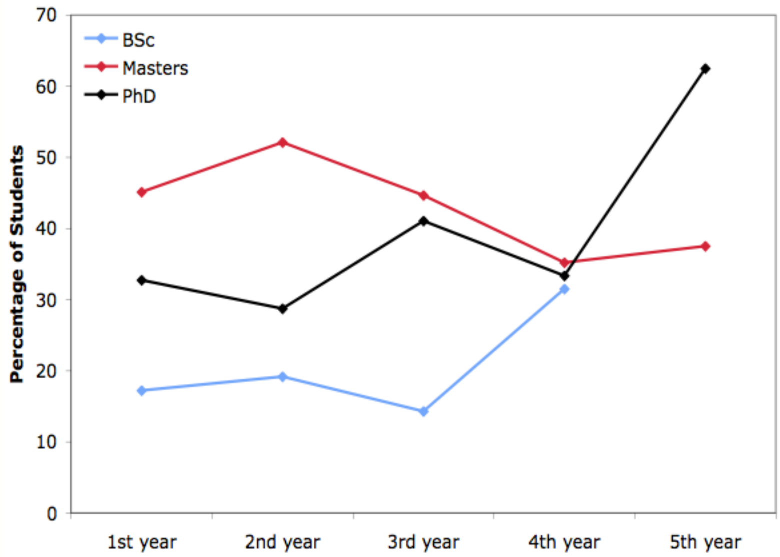


Figure 7.11: Intended exit point of undergraduate students enrolled in each year of the undergraduate programme in 2010-11. N(first year)=113, N(second year)=94, N(third year)=56, N(fourth year)=54 and N(fifth year)=24.

This level of interest from the cohort as a whole remained high after the transition into second year, with only a slight decrease in the number of students intending to do a PhD, although the proportion of the male cohort choosing this option (37%) was much higher than of the female cohort (17%).

By the start of junior honours, the percentage of students expressing an intention to do a PhD increased to 41%. In contrast, the distribution of intended exit points changed dramatically by senior honours. At this point in time there was a very balanced proportion of students considering BSc, Masters and PhD options. The physics degree programme at the University of Edinburgh has the option for students to complete a fifth year of undergraduate study for an Integrated Masters. For students who remain for the Integrated Masters, a year with a large focus on a Masters research project (worth 40% of the year), a PhD becomes an even more popular option. This data is however only a cross-sectional representation of each year in the undergraduate degree programme and is reliant on the response rate from students, which means that conclusive trends cannot be established from the data. The small number of female students in each cohort, particularly in fourth and fifth year, makes it difficult to draw definitive patterns of gender intentions from the data.

7.5 Students who change degree courses

When considering the issue of retention, this thesis investigated why some students make a conscious decision to swap degree programmes during their undergraduate career. A cross-sectional study of the 2010-11 undergraduate population illustrated that students made use of the flexibility of the system to change degree subject and/or endpoint over the course of their time at university. In some cases students changed their degree course on multiple occasions. These changes took place both within the range of Physics degree programmes and to non-physics degree courses such as Chemistry and Engineering.

7.5.1 Survey results

Students in the second to final years of the undergraduate physics degree programme were surveyed to determine the extent to which students utilise the flexibility of the Scottish degree system which enables students to transfer between degree courses, particularly during their pre-honours studies.

Of the 129 students surveyed during a core second year undergraduate physics course, 27 students (21%) commented on having changed their degree course by either subject or endpoint (BSc or MPhys) since entering first year, with a further 6 students commenting that they intended to change within the next academic year. Examining this for each gender, the proportion of the female cohort (24%) who had changed their degree subject and/or endpoint since first year was found to be slightly higher compared to the proportion of males (20%). This increased to 34% of the total surveyed population in third year and 41% in fourth year. In both third and fourth year the proportion of male students was much higher than that for females (44% and 45% for males and 7% and 25% for females). In the third year population surveyed only one female student indicated that they had changed their degree course. The relatively small sample size of female students in both year groups makes it difficult to gain a reliable conclusion from these results. Nevertheless, the overall findings suggest that students make good use of the flexibility inherent in the degree programme structure.

The subjects from which students changed were also examined in more detail. As mentioned previously, in second year 27 students changed their degree intention. Of these students 11 (7 male and 4 female) changed from a non-physics degree course, such as Chemistry or Engineering into Physics. Changes within the physics department were also very popular, with second year students choosing to change to a different physics degree programme (for example changing from an MPhys Mathematical Physics degree to a MPhys Physics degree). The remaining surveyed students changed from physics into the School of Geoscience.

By the time students reach their third year of studies there are no longer any students taking physics as an outside course. Of the 19 students who indicated that they had changed their degree course, 15 had made this change within the School of Physics. Two male students began their university career as Chemistry or Engineering students and one male and one female student had entered from the School of Mathematics. The number of fourth year students who had begun their undergraduate studies outside the physics department was again very small. Only 4 of the 18 students who had changed their degree course had converted from a non-physics degree. All integrated masters students had entered university with the intention of completing a physics degree.

7.5.2 Students who transferred out of the physics degree programme

When looking at the progression of students from one year of the physics degree programme to another, it is unclear from quantitative data when within that year the decision to withdraw or transfer from the degree programme is made. Despite approximately 40-50% of students enrolling in first year of undergraduate study not completing a physics BSc or MPhys degree [97], only the destinations of students who have transferred to a different degree programme within the University of Edinburgh are known. The destinations of those students who withdraw from the university or transfer to a different university are unclear.

In 2011, 35 students (20 males and 15 females) who transferred out of a physics degree programme to pursue studies in other disciplines at the University of Edinburgh between 2006-10, were contacted and surveyed about their reasons for choosing to study physics initially and their reasons for ultimately leaving the programme. Of these students, 17 responded to the survey (7 males and 10 females). As part of this process, students were asked to think back to the time just before they left the physics programme and answer seven questions on a 5 point Likert scale (ranging from Strongly Agree to Strongly Disagree). These responses were then collapsed into three categories: Agree, Neutral and Disagree. Male and female responses to each of these statements can be seen in Table 7.12.

Table 7.12: Number of male and female students who replied Agree (A), Neutral (N) and Disagree (D) to each statement on the survey regarding transferring out of degree programme. Percentages correspond to the percentage of male and female cohorts respectively.

	Males			Females		
	A	N	D	A	N	D
The subject was not what I had expected it to be	4 57%	0 0%	3 43%	3 30%	1 10%	6 60%
I did not perform as well in my exam as I had hoped	2 29%	1 14%	3 57%	2 20%	2 20%	6 60%
I applied myself to my studies to the best of my ability	5 71%	2 29%	0 0%	6 60%	1 10%	3 30%
I could do the physics but had problems with the maths	1 14%	2 29%	4 57%	1 10%	0 0%	9 90%
I had significant personal and/or health problems that affected my studies	1 14%	1 14%	5 71%	3 30%	2 20%	5 50%
I became more interested in my outside courses	5 71%	1 14%	1 14%	8 80%	1 10%	1 10%
I found my Director of Studies in physics helpful and supportive	5 71%	1 14%	1 14%	9 90%	0 0%	1 10%

A chi-squared test of each distribution showed no statistically significant differences at the 95% level between male and female responses to any of the questions. When asked if the subject was different to their expectations, 57% of males agreed, compared to 30% of the female population. This implies that when choosing university courses students may not have sufficient awareness of the course content and what the subject entails. There may exist a discrepancy between a student's expectations of the subject and the prescribed course outcomes or objectives.

It was not the case that students made the decision to transfer to another degree subject as a consequence of their course performance. Only two males and two females answered either 'Strongly Agree' or 'Agree' to the second statement, with the majority of students disagreeing. This is also reflected in the high proportion of responses from both genders who answered favourably to the statement "*I applied myself to my studies to the best of my ability*".

It was hypothesised that a large proportion of students who transfer out of physics do so because they have been discouraged by the number of compulsory mathematics courses involved in the physics degree programme. This was not reflected in students' responses. Only two students indicated any problem with the mathematics. When asked if they had found their Director of Studies helpful and supportive, 14 of the 17 students who responded to the survey replied favourably. Results suggest that students were not transferring from the physics programme as a result of problems with the physics course but because they became more interested in other courses taken in their first year of study. Overall, 71% of males and 80% of females indicated that they had become more interested in their outside courses. Only two students indicated that this was not the case. This result was explored further through interviews with a selection of survey respondents and will be discussed in the next section.

7.5.3 Discussion of qualitative survey responses

Students who had transferred from the physics department into other degree courses within the University of Edinburgh were approached to discuss, through qualitative interviews, the reasons for their decision to choose to study physics and for subsequently transferring to another degree. These interviews aimed to expand on the previously collected quantitative survey responses discussed in section 7.5.2. As in the interviews discussed in Chapter 5, these interviews were recorded using a Livescribe pen [150]. A full transcript of each interview was made. Seven students (3 males and 4 females) participated in the study. Results from these interviews will be presented alongside comments left by students who completed the online survey discussed in the previous section.

At the start of each interview participants were asked to discuss where their interest in physics originated; whether at school or from personal experiences. Several students commented on the impact of specific experiments or demonstrations they had witnessed. While the majority of the interviewees stated that their decision to continue with physics started in the later years of secondary school, while studying for their Advanced Higher or A-level exams, one female recalled being fascinated by a demonstration of centripetal force (a ball being swung in a circle above the teacher's head) and the fact that this led her to begin reading popular science books at a young age. The popularity of experimental physics was a key feature of several interviews. A male student voiced his disappointment that:

> *"we had to wait until second semester to do that* [practicals] *so I was a little disappointed with that. I was also not a fan of the maths which was more tutorials and lectures. Expected to be more hands on."*

The use of practicals or lecture demonstrations allows students to see the real life applications of physics concepts learnt throughout the course. The desire for earlier opportunities to take part in practical physics courses was not a view shared by all students. Some students who replied to the online survey indicated that they had chosen to transfer to a Mathematics degree. A common theme in their comments was their increased interest in the theoretical aspects of the subject.

When asked about the transition from school to university, the students who had completed Advanced Higher or A-level physics felt that the transition was easier than expected, with several students attributing this to the overlap of material between the school curriculum and the first year course content. Because of the range of educational backgrounds of incoming students discussed earlier in this chapter, the first year physics course is targeted to cover material accessible to students who enter with different qualifications and ensure that all students have the same content knowledge at the end of their first year of study. A consequence of this is that, particularly in first semester, some of the concepts discussed in lectures and tutorials have already been introduced in school courses, although the teaching methods and mathematical content may differ. One female student commented that:

> *"I think it was less of a jump than I was expecting, but that might be due to the fact that I did Advanced Higher physics in high school and I think it is kind of similar to first year university physics ... yeah there was quite an overlap. So in terms of learning and how you learn it was different but I didn't feel it was a big jump, like the biggest jump was more probably in the social aspect of university and in living away from home was maybe a bigger impact."*

This view was not shared by all students. Those who had left school with a Higher physics or maths qualification stated that they found it a larger transition than expected and that it was the style and rate of teaching that they found most difficult. One male participant who took Higher physics stated that:

"I think it was kind of too big a step almost because we suddenly had to do twice as much maths as I was really used to. And at school I did just Higher maths. I didn't do Advanced Higher ... I just found it quite difficult to suddenly be doing maths which was really aimed at a broader spectrum for Chemistry and Physics, and it just seemed a bit too much as well."

When asked if this was a major contributing factor to his decision to convert to a degree in Biological Sciences, he agreed.

The School of Physics and Astronomy offers entrants the option of enrolling directly into second year. Those that choose to do this must achieve higher entrance requirements compared to those entering into first year. Two of the students who participated in these interviews entered through the direct entry programme. One female student, who was from an American school background, said that she chose to enter straight into second year due to financial reasons as an international student. She was happy with her choice as it allowed her to challenge herself and it was a bit more *"mathematically rigorous"*, which she enjoyed. Conversely, another female student wrote:

"I still wondered if I hadn't done direct entry whether I would still be studying physics."

She went on to comment that she struggled with the mathematics content of the second year course and felt that there needed to be further guidance with the maths that she missed from opting not to complete first year. The success of the direct entry programme is very dependent on the ability and dedication of individual students.

In order to further understand the high proportion of surveyed students who stated that they became more interested in their outside courses, discussed in Section 7.5.2, the interviewer asked participants if they took advantage of the Scottish system of being able to choose a third of their first year courses from outside their degree programme. One male student, who took Medical Biology as an outside course in his first semester, said:

"I took Medical Biology in second semester because I wanted to see what the biology courses were like. Well, it was really just to see whether I really wanted to be changing courses at that point. So I went for it for a couple of weeks and I really enjoyed it and then asked to change at the end of first year."

It was evident from discussions with students that the outside courses chosen in first year enabled then to explore other subjects that were not available in secondary school curricula, for example engineering. By taking them as an elective they were able to use this year to look into other subjects that might be of interest to them, whilst still working towards their chosen degree. One student commented that their Director of Studies had specifically recommended that they choose their outside course from something they were interested in and may want to considering pursuing later on. One key feature of the discussions was that students transferring to Engineering degrees often changed their degree course because they found engineering to be a more applicable subject and one that could offer more job opportunities after graduation. This was evident in the statement:

"I felt that none of the [physics] *subject matter was linked to real world. It was quite interesting, but I did not feel it was preparing me for any career path. I chose Mechanical Engineering as it uses the principles of physics to solve real world practical problems."*

A participant who transferred to the School of Engineering said that she found the theory hard to understand during the lectures and preferred discussing the applications of physics concepts. Another student commented:

"I felt that I had more career opportunities with Engineering and that studying an applied science would be more useful in a workplace than physics. I am very practical minded, and physics seemed like too much theory."

When considering the attrition rates for physics at the University of Edinburgh, it may be beneficial for instructors to highlight the range of potential job prospects available to physics students. By calling attention to career opportunities in the early years of the undergraduate degree, rather than focusing on them only at the end of the degree programme, those students who were previously unaware of the options may make the decision to stay on the degree course. Some students feel that the majority of physics graduates enter business after finishing their degree rather than continuing in science careers. One student exaggerated this in their comment left on the online survey which stated:

"80% of physics graduates go on to become investment bankers and I want to use my degree. Engineering had much better job prospects."

Students were very positive about the ease with which they were able to transfer from one degree programme to another. As in results from the online survey (Table 7.12), they were keen to comment on the positive guidance they were offered by their Directors of Studies.

The key points to be taken from this analysis are that, in general, students based their decisions on a greater interest in outside subjects taken over the course of their first year. The students who transfer out are not necessarily the weakest students, with the majority attaining A or B grades in their first semester physics examination, nor does there appear to be an intrinsic gender bias. For the majority of students, their reasons for choosing to study physics at university originated from their enjoyment of it at secondary level and the fact that it was one of their strongest subjects at (Advanced) Higher or A-level. This was consistent with results found from the surveys in Section 7.4.2. Many were determined to emphasise that their decision to transfer out of physics was not a consequence of a failure in the physics teaching, but instead a result of a change in their personal interest. This was made clear in a final statement by one student:'

"The change in course had nothing to do with the way Physics was taught. I thought the lectures were very engaging and the lecturers were extremely enthusiastic. It was merely a realization that I wasn't interested enough in the subject and I wanted to know how things were applied, instead of studying the reason for things happening."

7.6 Graduating students' perspectives on their degree experiences

Studies of student progression through the undergraduate degree programme have shown that a significant proportion of undergraduates leave the physics programme between first and fourth year. It has been found that there exists no difference in the proportion of male and female students leaving during these years. For example, 46% of female students and 45% of male students who started the programme in 2008-09 graduated from the School of Physics and Astronomy [97].

Considering the final degree classification of graduating students, it was found that male students were more likely to receive a first class degree classification on the MPhys programme than female students [97]. Between 2008-2013, 46% of males were awarded a first class MPhys degree, compared to only 18% of females. It needs to be noted that the number of female students in each year group remained very small, with just 33 females graduating over these five years, compared to 155 males. Looking at upper second degrees (2.1) awarded, females appeared disproportionately more likely to gain a 2.1 (64%) than males (33%). This was not the case for those graduating with a BSc degree. The proportions of male and females awarded a first class honours were very similar (28% for females and 26% for males) [97].

After graduating from university a large proportion of physics graduates do not stay in the field of academia, or even science, with many using their degree to enter careers in business. In 2011-12 and 2012-13 students in their final year of the BSc and MPhys degree programmes were surveyed about their experiences of studying physics in undergraduate courses, the results of which are discussed in this section. The Course Experience Questionnaire (CEQ), developed at Curtin University [183], was administered online using 'Survey Monkey'. In addition to this, paper copies were made available to students at a reception held for graduating students. A copy of this survey is shown in Appendix J. At the end of the survey additional questions relating to the degree programme on which they were enrolled (BSc or MPhys) and their future career intentions were asked.

7.6.1 Results of survey questions

The survey contains 23 statements to which students mark their level of agreement on a 5-point Likert scale. Chi-square tests carried out on the responses to each of these statements showed no statistically significant differences between male and female responses or between the 2010-11 and 2011–12 year groups. Consequently, the results presented in this section represent responses from the combined 2011-13 population. In total 84 students completed the survey. As well as the statements mentioned, students were asked two open-ended questions about their overall experiences of their degree:

1. *What aspects of your degree programme were most in need of improvement?*

2. *What were the best aspects of your degree programme?*

A selection of survey statements will be addressed in this section, alongside qualitative comments provided by students.

Statement 2: "The teaching staff normally gave me helpful feedback on how I was going."

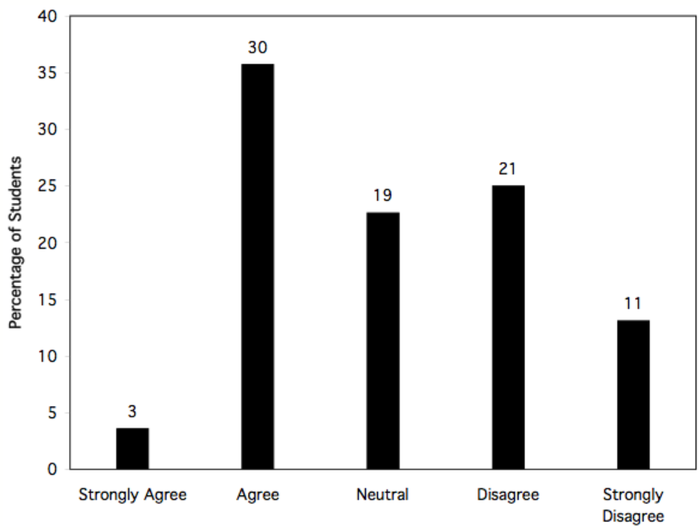


Figure 7.12: Response profile for 2011-13 students to Statement 2 of the graduating student survey. Numbers above each bar represent the number of students selecting each response. N=84.

The response profile of students' answers to this statement is shown in Figure 7.12. Although the most popular answer to this statement was 'Agree' (36%), a large number of students responded negatively. In addition to this, 23% of graduating students remained neutral. This indicated that students felt they may have benefited from additional feedback from teaching staff throughout their degree courses. One student's commented on this by saying:

> *"The quality of responses to submitted work and time given to individual interaction with lecturers/tutors were the poorest aspects of this course."*

The issue of feedback for written coursework assignments was commented on by many students and is widely recognised as a topic of contention in the National Student Survey [184]. This was seen as one of the few opportunities to get feedback on their individual progress and understanding during the academic year, which could then be used when preparing for end-of-course examinations. Alongside this, when asked what areas of the degree programme were most in need of improvement, many commented on the fact that the majority of junior and senior honours courses were assessed solely through a single examination. This led to students feeling that:

> *"it was difficult to gauge the progress we were making. It would help if the 4th year courses had hand-ins like the 3rd year ones."*

One female student said:

> *"I don't find grades given 100% based on exams very fair - I find myself, as well as many of my coursemates, learning a lot better from assignments, rather than cramming in all the course material in a week, and most likely forgetting it after a few months."*

This reflected evidence shown in Chapter 6, where female students were seen to consistently outperform male students in coursework assessments. The student above has commented on the fact that working through coursework may in fact improve students' retention of information, compared with short term learning for exams. This result was consistent with that found by Woodfield *et al.* in a survey of male and female undergraduate students, where students expressed that they found coursework to be a *"better test of their abilities and effort"* [68]. One student also expressed a wish to get more individual feedback on exam scripts.

Statement 7: "The programme sharpened my analytic skills."

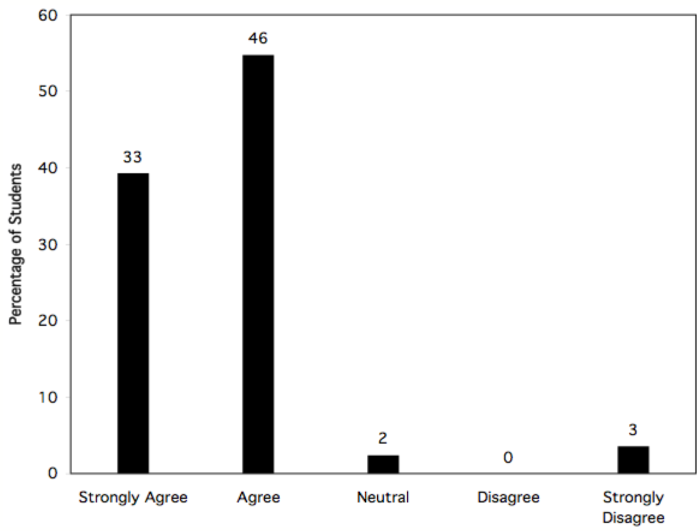


Figure 7.13: Response profile for 2011-13 students to Statement 7 of the graduating student survey. Numbers above each bar represent the number of students selecting each response. N=84.

When asked about the analytic skills learned during their undergraduate studies, students were very positive (Figure 7.13). Almost 95% of the survey population replied either 'Strongly Agree' or 'Agree' to Statement 7. This suggested that they were aware of the transferable skills gained during their studies. A female student stated:

"I have attained a very broad range of transferable skills which makes it easier to impress a prospective employer (i.e. in addition to problem solving, programming, mathematics and analytical skills, also obtained skills in team work, communication and presentation skills)."

Statement 11: "The programme developed my problem-solving skills."

Students responded similarly to Statement 11, shown in Figure 7.14, which asked students if they felt there had been an improvement in their overall problem-solving ability. A total of 88% of students agreed with this statement, with only 4 students (5%) indicating disagreement. This result is very encouraging. One of the primary aims of the physics curriculum is the ability for students to apply different concepts and problem solving methods to a variety of situations. This was acknowledged by a male student who wrote:

"The exams are too short. They can only test a limited part of the understanding since there is no time for effective problem-solving in 2 hours. And problem solving is the only way to test the understanding of students. Therefore, exams should be more difficult but should allow enough time for thinking and re-thinking over the problems."

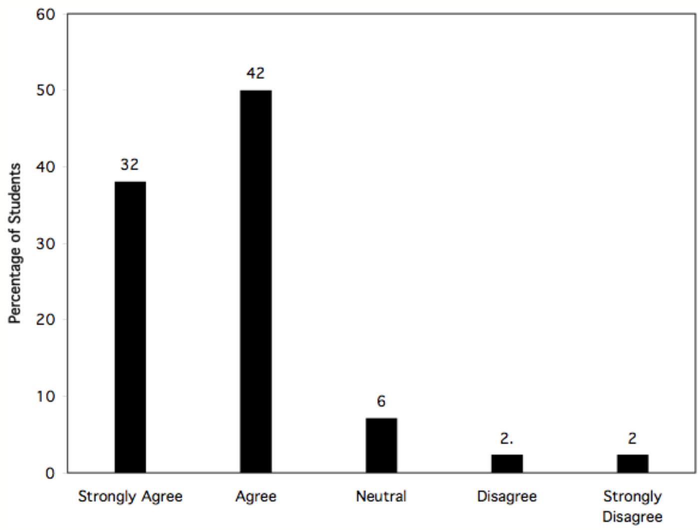


Figure 7.14: Response profile for 2011-13 students to Statement 11 of the graduating student survey. Numbers above each bar represent the number of students selecting each response. N=84.

Students' desire to improve their problem solving skills and skills that would make them *"more marketable"* for future employers was evident in the qualitative comments left in the online survey.

"... more of these types of courses where students are required to develop a core toolset for solving problems ... I would advocate open book exams or other forms of continuous assessment where the student has full access to resources ... That way the problems given would not be ones that could simply be looked up, but the student actually has to use his problem solving facilities."

Open book exams have been introduced in the first year courses. Despite students being able to bring course notes into the exam, it has not led to a significant increase in final marks, but has instead led to questions involving more application of concepts and problem-solving, rather than the recall of information. Graduating students' responses suggest that introducing such a change, or a greater emphasis on coursework assessment, in later years could be beneficial and promote more productive study techniques.

Statement 20 "My degree programme helped me to develop the ability to plan my own work."

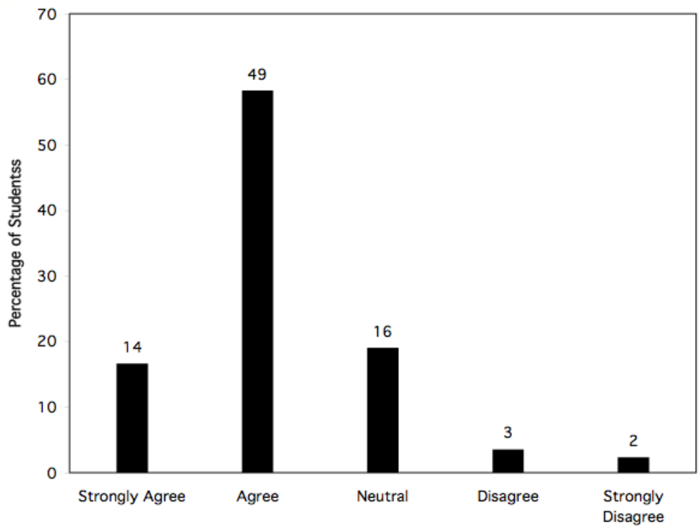


Figure 7.15: Response profile for 2011-13 students to Statement 20 of the graduating student survey. Numbers above each bar represent the number of students selecting each response. N=84.

Another important aim of any degree programme is to improve students' independent learning. This is a particularly important skill for those intent on postgraduate study. In response to Question 20 of the survey, 75% of students answered either 'Strongly Agree' or 'Agree', as shown in Figure 7.15. Only 6% of students answered either 'Strongly Disagree' or 'Disagree'. Several of the survey responses came from students on the Integrated Masters programme. These students stated that one of the best aspects of their degree programme was their Masters project, upon which their final year of study is focused. Students commented that it was *"an immensely rewarding experience"* and that it *"taught me how to be fully self-reliant"*. The aim of the masters project is to give students an opportunity to cultivate both technical and presentation skills which will be useful, both if they decide to remain in a research-based environment or choose to leave academia.

In addition to the independent research skills developed over the course of the degree and final year research projects, students commented on the insight into *"real research"* these provided. They commented that they enjoyed interacting with staff about their personal research areas:

"Amazing is the way that lecturers and the environment pushes you to study and learn. It stimulates the mind and definitely it teaches you to have an intuitive mind for many problems, even not related with physics. I really do feel I have a different way of viewing the world."

This final year acts as a time in which several students make the decision to either remain in further education and pursue postgraduate qualifications or to seek employment after graduation. One male student stated that, after completing his fifth year research project:

"I have since gained the inspiration to pursue further education from that ... from having enjoyed the projects so much."

Statement 23 "Overall I was satisfied with the quality of this programme."

The final survey statement asked students to indicate their overall satisfaction with their experience of the physics degree programme. The response profile for this question is shown in Figure 7.16. The results were very positive, with 71% of respondents choosing to 'Strongly Agree' or 'Agree' to the statement. Despite this, 20% of students indicated that they were not completely satisfied with the quality of the programme, indicating that there are areas in which students' comments could be taken into consideration to improve the overall degree experience.

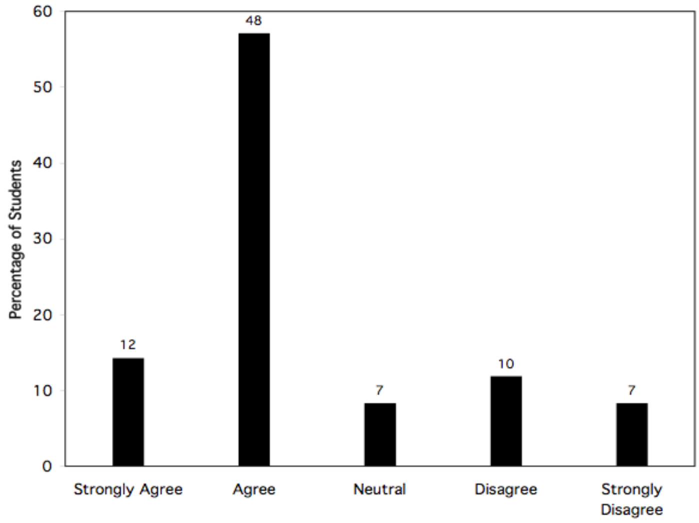


Figure 7.16: Response profile for 2011-13 students to Statement 23 of the graduating student survey. N=84.

7.6.2 Graduating students' intentions

The career intentions of graduating students were tabulated and are shown in Table 7.13. A total of 76 students responded between 2011-13. Of these students, 64 were male and 12 were female. With respect to exit points, 31 responses came from students graduating from the BSc programme and 45 from the MPhys programme. Students were asked if they intended to continue with postgraduate studies or if they planned to enter employment after graduation. Whether they had already secured a position for the following year was also noted.

Table 7.13: Future intentions of 2011-13 BSc and MPhys students. N=76, N(males)=64 and N(females)=12.

	All Students		Male Students		Female Students	
	BSc	MPhys	BSc	MPhys	BSc	MPhys
Postgrad secured	12	19	9	16	3	3
Postgrad intended	5	8	5	6	0	2
Employment secured	6	7	4	7	2	0
Employment intended	4	8	4	7	0	1
Undecided	4	3	4	2	0	1

It can be clearly seen that the majority of students surveyed (58%) intended to continue studying for a postgraduate qualification. This percentage was slightly higher for MPhys students (60%) than for BSc students (55%). Of those intending to pursue a postgraduate degree, a large proportion indicated that they had already secured a position, either here or at another university. Nevertheless, 30% chose ‘Postgrad intended’. A further 33% of students stated they were planning on entering employment after graduation, of which 52% had already secured a position. Finally, 7 students (9%) remained undecided.

When investigating potential gender differences in the intentions of physics students after graduation, it was difficult to draw clear conclusions due to the very small numbers of female respondents to the survey. Only 12 females responded, of which 8 (67%) were intending to go into a postgraduate position. The number of females who had already secured this position was equal for BSc and MPhys graduates. Of the remaining four students, two had secured employment, one intended to find a job outside academia and one remained undecided. A large proportion of male students intended to complete a postgraduate qualification (56%). This was particularly the case for those who completed the MPhys programme. A further 34% were intending to enter employment, half of which had already secured a position.

7.7 Chapter discussion and summary

In this chapter the attitudes and opinions of physics students were investigated at two keys points in their undergraduate education: at point of entry to first year and just before their graduation. Both quantitative surveys and qualitative analysis have been employed.

Attitudes of first year undergraduate physics students, both prior to and post teaching, were examined using the CLASS survey instrument. Students at the University of Edinburgh entered the degree programme with high levels of expert-like thinking compared to previously published results [84, 85]. Post-tests results, however, differed depending on the academic year in question. For example, students in 2010-11 showed a significant drop in their level of agreement with the expert response. This drop has been widely reported in the literature, although the reasons for this observed change in attitudes and beliefs are not conclusive.

Results collected from the two following years indicated no such drop in expert-like thinking amongst our first year students. The timing of this lack of change coincided with the introduction of the inverted classroom format in first year lectures [63]. The extent of the effect of this transition on students attitudes towards learning and studying physics is unclear, but may allude to the idea that introducing more interactive engagement methods into lectures has minimised the drop in students' attitudes.

Despite major students having a higher percentage expert-like thinking score than non-majors, there existed no statistical difference between the cohorts at the start of the first semester. Differences were found between the two cohorts in 2010-11 after two semesters. Non-major students showed a significant drop in favourable expert responses and the gap between majors and non-majors increased significantly by the end of the year. These observed differences cannot be explained by differences between entry qualifications (both populations achieving the entry qualifications for the physics degree programme) or prior learning for these two groups. One possible explanation may be a difference in major and non-major students' motivation, depending on the importance they place on courses that are compulsory for their degree programme and those that are chosen electives.

When investigating gender differences between students, it was found that in each of the three years analysed there were no differences between male and female students in either the pre- or post-instruction survey results. Both groups of students showed little change over the two semesters. This result differs from those collected at the University of Edinburgh in years prior to 2010, where female students showed significantly less expert-like levels of thinking after two semesters of teaching [177].

Interesting differences were observed when examining scores on individual categories. Students scored very highly on some categories, such as *'Sense Making/Effort'*, while they scored much lower on others, including *'Applied Conceptual Understanding'*. Of particular interest was the consistent increase in expert-like thinking scores by all year groups in the *'Real World Connections'* category. Changes in category scores must be dealt with with some degree of caution. Each category contains only a small number of statements. Therefore, small shifts in responses to individual statements can result in larger overall changes in category scores.

It is difficult to comment decisively on the reasons for these changes, but several factors should be taken into consideration when considering the presented results. There were differences in the way in which the CLASS survey was administered to students in Edinburgh compared to how it was presented to students by the original survey authors. First year physics students in Edinburgh were presented with the survey during a tutorial and took an average of 5-10 minutes to complete it. When the survey was first tested in the US students received it in an online format and were allowed three to seven days for its completion. Unlike in this study, students were also given course credit for participation. There are also reasons to suggest that the 2010-11 academic year had a different student population. There was a higher proportion of non-major students than physics majors and the proportion of females was dramatically higher than in proceding years. As has been mentioned previously, this year also marked the beginning of a selection process for choosing incoming students.

The second study presented in this chapter was a comparison of US and UK academics' attitudes, as measured by the CLASS survey. At the time of publication of this thesis no other studies have been undertaken to derive a level of expert-like thinking from physics academics in the UK. Results from members of the Institute of Physics showed a statistical difference between the opinion of male and female academics. Interestingly, female academics showed a higher level of agreement with the US experts. In fact, female academics had a higher expert agreement on all eight of the categories and this gender gap was significant for the *'Real World Connection'* category. Perhaps most noteworthy is the fact that five survey statements showed inconsistencies in the level of consensus amongst academics. Although the level of agreement with the predefined expert response from the original survey did not fall below 50%, the low agreement from UK academics raises the question of whether there are differences between experts' attitudes in the US and UK. The CLASS survey was originally compiled from responses from all male physics faculty members. Potential gender differences may need to be considered when analysing the level of favourable responses of different populations.

It is clear from first year survey responses that students enrol onto the introductory physics courses from a wide range of physics and mathematics backgrounds. A large number of students leave physics courses between the first and final years of the degree programme. What was of particular interest was the school leaving qualifications of students as they progress through the degree programme. Although A-level, Advanced Higher and Higher physics qualifications were all relatively popular amongst students in their first year, the number of students with a Higher physics qualification showed a decline towards the final years of the degree. Student intended exit points also changed over the five years of undergraduate physics. The majority of first year students intended to complete either a Masters or PhD qualification. While this intention remained relatively stable during pre-honours, the number of students considering leaving after the BSc increased greatly at the beginning of fourth year. This may suggest that students' experiences in junior honours may have had an effect of their desire to continue studying. Equal percentages of students intended to leave with a BSc, Masters or PhD qualification at this point in their degree.

Survey responses about students' choices to transfer out of the physics degree programme were very positive about their experiences in the department. It was clear from qualitative interviews that students based their decision on the fact that they had become more interested in another subject, often the subject chosen as an elective in first year. When asked about their experiences of their physics courses, many emphasised that it was not a failure of the physics curriculum or teaching, but simply a change in their personal interests. In terms of gender, there were no observable differences in male and female students' opinions. This may partially be the result of the small sample size, making it difficult to infer any gender differences.

Graduating students, surveyed just after their final degree examinations, highlighted several features of the degree programme they felt could be improved to enhance the undergraduate experience. An increase in individual feedback from lecturers, both on coursework assignments and exams, was a key issue raised by students. In addition to this, a female student stated that coursework provided a better learning opportunity compared to short term memorization that often occurred before exams. Students were keen to comment on the fact that they felt the degree improved their problem solving and analytical skills, particularly in final year research projects. Overall, students showed a high satisfaction in their undergraduate experience. As with the transferring students, the qualitative analysis showed no differences in male and female responses to survey statements or to the comments left by graduating students about their degree experiences.

Chapter 8

Conclusions and Future Work

8.1 Summary

This thesis has probed the existence and extent of gender differences in undergraduate physics courses in many different contexts. The gender issue has been explored, first from the perspective of male and female participation rates, secondly in terms of student performance in different courses and on different forms of assessment, and, finally, through students' perceptions and attitudes to learning physics and their experiences of the undergraduate degree programme. In this chapter a summary of the main findings of this thesis will be presented as well as suggestions for possible areas for future research.

Examining differences between male and female participation rates, both at secondary school and undergraduate level, has shown that females are consistently under-represented in physics. While the numbers of students deciding to study physics in their final year of secondary school and at university have shown an increase in the last decade, as discussed in Chapter 1, the proportion of female students has remained relatively constant. The proportions of females taking physics at Higher and A-level is approximately equal to that at undergraduate level. This suggests that it may be unrealistic to expect changes to be made to the gender participation discrepancy in physics once students have enrolled in higher education, and that we need to target this gender gap much earlier in students' schooling.

We have seen that the introduction of new teaching methods can have a positive effect on student learning, for example the use of interactive engagement techniques such as Peer Instruction or the 'flipped classroom' approach in lectures. In some cases this has resulted in a decrease in the gender performance gap. Consequently, it is important to continue to encourage instructors to try new teaching and assessment methods in their courses. Having identified areas in the undergraduate physics degree curriculum at the University of Edinburgh which have shown performance and attitudinal differences between males and females, it is important to consider reasons behind these discrepancies, as well as the options available to change the current situation.

8.1.1 Student participation

The undergraduate population at the University of Edinburgh discussed throughout this thesis is typical of UK physics departments in that female students are highly under-represented compared to males. Although the first year undergraduate cohort has had an average of 24% female students over the last seven years, direct comparisons between cohorts were complicated by variations in the number of females in each year group. The percentage of females enrolled in first year courses at Edinburgh is in fact slightly higher than the UK national average which has remained approximately 19% over the last decade [21]. Approximately 20% of students studying Advanced Higher and A-level physics in the UK are female. The fact that the proportion of female students studying Advanced Higher and A-level physics courses is approximately the same as that for undergraduate courses [14, 185], suggests that those students who want to continue with physics in higher education have the opportunity open to them and that university courses are maximising on their intake.

There existed no clear distinction between male and female students' reasons for choosing to study physics at university level at Edinburgh. Interviews with students who transferred out of the physics degree programme in their first two years of university emphasised that their decisions were based on a change in their personal interests and career intentions, rather than a negative effect of the physics programme. These results suggest that the participation gender gap originates earlier on and needs to be targeted at a time prior to future study decisions being made by school pupils, even as early as primary school [186]. It is therefore unrealistic to expect for there to be any effect on the participation gap once students have enrolled on university courses. Nevertheless, it is vital that students' interest and enthusiasm in the subject be maintained, alongside the promotion of career opportunities. This should help encourage students to remain in STEM careers after graduation.

8.1.2 Student performance

One of the primary focuses of this thesis has been on students' conceptual understanding of key physics concepts. This has been probed using both quantitative and qualitative methods, the results of which are discussed in Chapters 3 to 5. At the point of entry to the degree programme students showed a wide range of levels of understanding of Newtonian mechanics, as measured by pre-test FCI scores. After being exposed to eleven weeks of teaching, using interactive engagement techniques in both lectures and workshops, post-test results showed marked improvement by both genders. Results from seven consecutive diets of the first year undergraduate physics course have provided evidence of positive learning gains for the whole cohort, which remained high after the introduction of the 'flipped-classroom'. This is an encouraging result and demonstrates the effectiveness of introducing new teaching methodologies.

Investigating gender differences on entry to the degree programme found that male students had higher levels of conceptual understanding than females, with all year groups analysed showing a gender gap on the FCI diagnostic test. Despite the class as a whole showing good improvement at the end of the course, there remains some concern over the persistence of a gender performance gap after a semester of teaching. Although the gender difference was reduced, there continued to be a statistically significant difference between male and female scores in all year groups. Ideally, both cohorts would show equal levels of understanding of these fundamental physics concepts, as was seen by the complete closure of the gender gap at Harvard [55], however this result was not replicated by the data presented in this thesis. Nevertheless, the decline in the gender gap was consistent with previously published literature [56, 57, 62].

Perhaps even more compelling evidence for the existence of the gender disparity in Newtonian mechanics was the distribution of male and female populations across FCI performance quartiles. In many year groups approximately half of the entire female population was in the lowest performing quartile. It was also observed that a large proportion of students who began the academic year in this pre-test quartile remained in the lowest quartile at the end of the semester. This was particularly the case for female students. Perhaps additional support needs to be provided specifically to these students, whether that takes the form of additional help with regards to the course content or, more generally, to study skill strategies. Conducting this analysis consistently across three different UK universities [118], each with a different population of incoming students, illustrated that the gender performance issue is not unique to the University of Edinburgh, but is indicative of a wider problem as has been previously suggested by published results from US institutions [55, 56, 62].

A comparison of physics majors and non-majors indicated that gender was not the only differentiating factor in performance. Despite both majors and non-majors meeting all the necessary entry requirements for the first year course, there existed a difference between the initial levels of conceptual understanding of force and motion between the two populations. Those students who had the intention of completing a physics degree scored significantly higher than non-major students at the start of the semester. These differences may suggest that a student's perception of their studies may influence their performance depending on whether it is their main subject area or whether they approach it as a subsidiary subject. The motivation of students toward their studies of an outside course may affect their performance. The type of motivation is also an important factor to be considered [129]. Whether it is 'intrinsic' motivation, stemming from how interesting or satisfying a student finds the course material, or 'extrinsic' motivation, based primarily on achieving a specific outcome, may influence a student's course performance [129]. In this respect, non-majors' under-performance compared to majors may suggest that they choose to focus the majority of the time spent on their studies on their core degree subjects and perceive their performance on an 'outside' elective to be of lesser importance.

The identification of common misconceptions surrounding a physics concept amongst undergraduates is key to understanding where emphasis should be placed when addressing these concepts during instruction, both for the benefit of the class as a whole and for addressing the existing gender gap. Furthermore, students may answer questions correctly without a full understanding of the concept, a point made clear from student interviews presented in Chapter 5.

Understanding students' preconceptions is also important for designing effective physics problems and multiple choice distractors for future concept evaluation instruments. Misconceptions emanate from students' prior instruction and learning and it is widely recognised that, particularly for Newtonian mechanics, students' misconceptions are often the result of their personal experiences in the physical world [131, 132]. There is extensive literature discussing the origins of these beliefs and how, through instruction, they can develop into a more expert understanding [187, 188]. Theories suggest that these perceptions of the physical world can develop in early childhood and students bring these knowledge schemes with them to future learning situations [188].

As well as the need to fully understand what the misconceptions held by students are in order for teaching to be used to target individual preconceptions, in some cases it needs to be noted that the difficulty lies in individual student's struggles to resolve their knowledge of the physics concept with their 'common-sense belief' or their view of the 'real world' [132]. For example, in many cases when students were interviewed they were able to correctly state Newton's laws, but had difficulty applying these laws to physics problems or reconciling them with their own personal experiences. This barrier to learning new concepts is not restricted to physics education, but is applicable across all disciplines and is particularly relevant as students make the transition from school to university. This is a period where students are exposed to, not only new material, but new teaching methods and are required to undertake more independent study. In some instances the necessary conceptual change requires students to add new information to an already existing knowledge structure [130, 132, 189]. In this case, students try to form connections between this material and that previously studied. Alternatively, students may need to revise their already existing beliefs [189]. Perry's model of intellectual development describes the transition of students' development through their university career [190]. He suggests that most students enter university with an assumption that knowledge is certain and that there is only one correct answer which is handed down by a figure of authority ('dualism'). As students progress through the course they become more aware that knowledge is relative and context dependent, and are therefore better able to resolve the discontinuity in their understanding. Tabor discusses the consequence of the presence of misconceptions by students on their learning and the need for instructors to explicitly confront incorrect preconceptions [191]. He comments that if students are presented with new material which is inconsistent with their prior intuition, it is important for instructors to challenge students' understanding and make the distinctions between their conflicting ideas clear. Similarly, steps must be taken if students are unaware of the relevance of concepts that they have already learnt to those being currently presented.

Smith *et al* relate the need to overcome these misconceptions as a transition from naive theories towards expert concepts, and comments that learning involves both the "*acquisition of expert concepts and the dispelling of misconceptions*" [132]. This is similar to the discussion by Chi *et al* on the difference between how novices and experts categorise physics problems, whether by physics principles or surface features [162]. This was observed with students from Edinburgh, Hull and Manchester, who showed evidence of trying to solve problems by associating them with previously answered questions set in similar contexts, rather than by identifying the relevant fundamental physics principle. Students may base their choices on their own experiences, whilst 'experts' in a discipline often base their solutions on a conceptual model. However, it may not be the case that all misconceptions arise through students' personal experiences and some students' confusion may be a result of the presentation of material by instructors or textbooks [192]. Because instructors know what the underlying meaning of a statement is, they may be unconscious of where the confusion arises and of the need to explicitly clarify the correct interpretation of a concept.

The results from the exploration of multiple choice responses to force and motion questions indicated the prevalence of common misconceptions amongst our first year students, and observations and interviews with students were in agreement with the literature discussed above, with some participants referring to their personal experiences. Furthermore, the results of this thesis imply that there may be an additional gender dimension to the extent to which students hold on to these misconceptions. Analysis of the specific incorrect multiple choice answers chosen by students on the FCI revealed that males and females may hold on to certain misconceptions to a different degree, as was shown in Chapter 4. The qualitative results reported in Chapter 5 supported these quantitative findings and pinpointed several weaknesses in students' understanding of key physics principles. For example, one prevalent misconception noted was students', particularly females', incorrect assumption that all motion required the presence of a force and that the magnitude of this force dissipates over time if the source is no longer in contact with the object [160, 161]. This misunderstanding of 'impetus', which persisted throughout several test questions, was observed when answering a question involving throwing an object upwards. For example, females were more likely to assume that there exists an upwards force on the object after it has left the thrower's hand. As has been well documented in the literature, results also showed that students had great difficulty when applying Newton's Third Law to a situation [136, 137], although no gender discrepancies existed, with both cohorts showing equal levels of confusion. These results, along with other preconceptions discussed in Chapters 4 and 5, emphasise the importance of addressing such gaps in understanding in order to lay the foundations for the assimilation of future knowledge. Why certain misconceptions are more apparent in females is still unclear. One possible explanation may be that females have more difficulty abandoning their previous convictions than males, and are subsequently more likely to maintain misconceptions even after being presented with new information which contradicts their prior knowledge. Conversely, it may suggest that males are less likely to try to relate different physics concepts to prior experiences.

Results from qualitative interviews at the Universities of Edinburgh, Hull and Manchester demonstrated no overarching differences in the approach or reasoning of male and female students. However, it should be noted that the sample size was small and it was difficult to recruit participants from the lower performance quartiles where the largest gender discrepancies in performance existed. The time scale involved in completing such a study for a large sample of students or for a whole cohort would be extensive, but such a study could allow for a better comparison of gender approaches to problem solving in addition to their conceptual understanding.

One emerging feature of the qualitative interview analysis was the difference in confidence levels of female and male students when indicating their chosen answers. Interestingly, females were more likely to show a higher degree of confidence when giving an incorrect response than males. Conversely, male students showed a higher confidence level in their correct answers than females. One could speculate that males are more conservative in showing their confidence when they are uncertain of the correct solution. This observed difference may also suggest that female students are more likely to choose an answer based on their first instinct, or one which coincides with their preconceptions, whilst male students may consider each possible answer option before stating their final response to the question. A further consideration is that those female students who showed very high levels of confidence in their chosen response may be doing so as a result of inherent stereotype threat and consequently feel they need to show a high degree of confidence in their work [30].

Multiple assessment methods have been used in the study of students' physics performance, each of which contributes a different perspective on the undergraduate gender issue. Whilst at university students' performance is measured both through continual assessments, such as weekly written assignments, and final examinations. A clear trend in female bias towards continuous assessments was witnessed. As well as in physics, this gender gap was seen in biology and chemistry courses at Edinburgh, both of which have much higher proportions of females in their undergraduate populations. The under-performance in coursework of males compared to females has been witnessed across science and social science disciplines, both at secondary school and university level [3, 56, 66, 68]. Some studies have also shown a contrasting gender gap in favour of males on examinations [3, 62]. In the case of undergraduate degree courses investigated in this thesis, no such gender bias was seen in examinations. Having recognised this consistent gender gap in assessment performance, we must ask why females do better on continual assessments than males. It has been debated the extent to which the gender discrepancy is caused by the structure of the assessment and how much is a result of students' perceptions and confidence level in their ability to perform well [69, 193]. It can be argued that coursework and examinations result in different types of learning and that instructional recommendations should be based on which offer the greatest learning benefits to students. Examinations can result in a focus on short term rote learning and memorisation, whereas continual assessments provoke students to consider individual concepts in more depth [68]. Comments from graduating students, presented in Chapter 7, discussed the issue of coursework and open-note exams, with many suggesting that these promote more productive learning than courses assessed purely through closed-book examinations. They felt that weekly written assignments allowed them to judge their performance at various stages throughout the semester. Unsurprisingly, in addition to this, students expressed a desire for an increase in coursework and examination feedback, a result widely recognised and documented by results from the National Student Survey [184]. Another possible explanation for this female bias in coursework scores may be a gender discrepancy in student diligence towards their studies or the amount of time they spend working on their written assessments. In light of the results presented in this thesis, in which females outperformed males in weekly assignments and laboratory assessments, it must be asked whether courses assessed entirely through coursework would preferentially favour females.

8.1.3 Student perceptions

The topic of students' perceptions to learning physics has been touched on in Chapter 7 of this thesis. The attitudes of students towards studying physics were measured using the CLASS survey for three first year undergraduate year groups. No gender differences were found in either the overall pre-instruction or post-instruction CLASS results for each of the three years analysed. In the first of these years students showed a decline in expert-like thinking, a result consistent both with previously published literature [84, 85] and data collected from first year students at the University of Edinburgh prior to 2010 [177]. In contrast however, one of the main findings of this research showed that in 2011-12 and 2012-13 there was no statistical decrease in the expert-like thinking of the first year cohort. This change coincided with a reconfiguration of the format of lectures to a 'flipped-classroom' approach [63]. The timing of this change presents the possibility that the increase in interactive engagement methods and peer discussion associated with the format of the 'flipped-classroom' may result in the drop in positive expert-like thinking being minimised. One could further speculate that by encouraging students to partake in peer discussion during lectures, students are encouraged to develop some of the characteristics of an 'expert', where an expert tends to reason through a problem qualitatively with reference to the relevant underlying physics principles [162]. If following year groups continue to show no drop in expert-like thinking after two semesters of teaching, this may provide additional evidence for the effectiveness of the inverted classroom approach in lectures and the benefit it has on maintaining students' expert-like attitudes to the subject.

In addition to overall favourable and unfavourable levels of agreement with the 'expert' responses to learning physics, students showed different levels of expert-like thinking on certain categories of the CLASS survey. In particular students showed a consistent increase in scores in the *'Real World Connections'* category between pre- and post-tests. Although the small number of statements within an individual category means that conclusions made between comparison of cohorts on a category basis must be treated with some caution, this high level of *'Real World Connections'* is consistent with results from first year diagnostic tests and interviews suggesting that students often try to contextualise new physics concepts within a situation with which they are personally familiar [188].

Results from the use of CLASS with the undergraduate population led to an investigation into the similarities and differences between 'expert' physicists in the UK and US. This is the first time that such a measurement of the attitudes of UK physics academics and industry members has been undertaken and, unlike in the original US survey validation, enabled a comparison of male and female academics to be made. The data collected indicated that the expert view of UK academics was below that of the 100% US academic level, suggesting that academics at UK institutions may have different viewpoints to those in North American universities. Some individual survey statements included in the CLASS survey did not show an overwhelming agreement with those of the US faculty members who helped in the survey validation process [84]. While the results from UK academics did not have the effect of changing the defined 'expert' response from 'Agree' to 'Disagree' (or vice versa) for any statement, the data and comments left by participants suggest that we should not expect a 100% agreement with the US experts for some of the statements. This raises the question of how expertise is defined and why there are differences between US and UK, and male and female academics. This also raises the question of what level of expertise should be expected of students. Results from this study may suggest that it is prudent to explain students' favourable agreement scores above a certain percentage level to be evidence of expert-like thinking, allowing for some natural variation in students' personal attitudes. Nevertheless, the use of the instrument as a measure of the change in student attitudes is unaffected. Because we compare students expert-like thinking scores prior to and post-instruction using the same instrument, any potential differences between expert opinions do not affect comparisons we make of the change in students' attitudes over the course of a period of teaching. As well as an overall statistical difference between male and female UK academics, gender discrepancies between academics were evident on several independent CLASS statements. These referred to how useful they felt physics knowledge was in their day-to-day life and how able they felt to transfer knowledge from one physics problem to another. In each case females showed a higher level of agreement with the US expert response.

The final topic of results presented in this thesis considered the opinions of students who chose to transfer out of physics to another degree programme and those of students preparing to graduate from the degree programme. Overall, students were very positive about their experience of the physics teaching, with students stating that their decision to leave the physics programme was based on their personal interest in another subject or future career prospects. Graduating students commented on the wide range of skills they acquired during their studies, particularly those used in final year research projects. Both of these qualitative studies showed no differences between male and female responses, suggesting male and female students share similar views on their degree experiences overall. These observations also compound the view that there needs to be more emphasis placed on encouraging the study of physics and the promotion of potential career opportunities in early years of education if we want to increase the number of physics graduates in years to come.

8.2 Future research

In order to improve the gender balance in participation in physics courses at university level, it is important to encourage more students, and particularly females, to study STEM subjects at an early age. It is sometimes questioned why we should aim for more gender equality in STEM subjects. Increasing the number of both male and female science students is important, for the growth of both industry and academic research and it is therefore important to ensure that efforts are made to make physics and other sciences available and approachable for women who want to study these disciplines.

Students who have enrolled on the first year introductory 'Physics 1A' course are in essence a self-selecting group who have overcome stereotypes and gender obstacles in order to choose to pursue their interest in the subject. Having found no gender differences in the reasons for first year physics students choosing to enrol on the physics courses, one interesting area for future research would be to ask pupils in secondary school, prior to the age where critical subject decisions are made, to comment on their reasons for continuing with physics (or equally their reasons for choosing not to continue studying physics) at tertiary level and their perceptions of physics as a career option. This may allow teachers and university instructors to better understand students' attitudes towards physics. One hypothesis is that students are unaware of the vast range of careers available to physics graduates. This was suggested by interviews with students who transferred out of the physics degree programme. They commented that they found other subjects to be more applied and offer "*much better job prospects*". The IOP 'Closing Doors' report suggests that gender stereotyping is an important factor and actively needs to be addressed, even at primary school [185]. If it was evident that these stereotypes continued to exist or that students at secondary school were unaware of the range of job opportunities available with a physics degree, greater emphasis could be placed on providing additional career advice at school level and at the point of entry to university and continuing into the early years of the degree programme. Looking at the destination of previous physics graduates and disseminating these results to current students may also broaden their outlook on future career options.

There are several areas in which further exploration into university physics performance could be undertaken. As noted from survey data presented in Chapter 7, students arrive at university having completed a variety of school qualifications. Although all enrolling students have achieved the necessary entrance qualifications for the physics degree programme, differences in the course syllabi of school courses means that it is not guaranteed that all first year students have covered the same material prior to university. This is borne out by the wide range of familiarity with basic concepts of Newtonian physics. It would be interesting to explore whether there was a correlation between students' school leaving qualifications (and, for A-levels, the specific physics modules completed) and their performance on both the introductory physics diagnostic test and degree assessments.

A significant portion of this research has concentrated on measuring students' conceptual understanding of Newtonian mechanics using the well established FCI diagnostic test. At the time of writing this thesis only two years of data pertaining to the FCI were available, with all previous year groups completing the FCI_{ext}. Despite this, the results from the 2011-12 and 2012-13 cohorts showed similar male and female differences to the prior five years. While in some cases several years of data have been amalgamated into a single data set, allowing for higher confidence in statistical results to be achieved, there nevertheless remain limitations in some of the conclusions that can be made due to the small number of participants in the populations explored. This is particularly the case when considering female cohorts. One primary example of this is in the longitudinal coursework and examination analysis conducted in Chapter 6. While this method had the benefit of eliminating the effect of differences between consecutive year groups of students, the number of students in the final years of the undergraduate degree programme is much smaller than those in pre-honours years, making gender analysis particularly difficult. Analysis of additional data sets in future years could provide further confirmation of the observed gender differences and any changes in the gender gap over time.

Question by question analysis and qualitative interviews with first year students have gone some way towards identifying and confirming specific common misconceptions relating to force and motion. While out of the scope of this project, finding the origins of such misconceptions, or the time in students' education when these misconceptions manifest themselves, could prove very useful in trying to eliminate the preconceptions which the students hold when entering university. Results from the FCI pre-test at Edinburgh suggest that such misconceptions exist when entering university and therefore originate during secondary school when students are first introduced to each topic. Administering the FCI, or a similar diagnostic, to students in their final years of secondary school could not only confirm the existence of the same misconceptions, but also be used to inform teachers of weaknesses in students' understanding at the point of initial introduction to the subject material. When the FCI was first used by Mazur to gauge the effect of Peer Instruction teaching on students at Harvard University, he felt confident in their ability to perform highly on the test questions [194]. Students' scores were surprisingly low compared to the perceived difficulty of the conceptual inventory test. It may therefore not necessarily be clear or obvious to instructors that their students are misunderstanding certain physics rules.

The fact that female students consistently outperform male students in continual assessment, both in physics and other STEM subjects, strongly suggests that there may be a gender bias in term of the type of assessment which students need to complete. What causes this gender bias in assessment type is unclear [68, 69, 193]. One hypothesis is that coursework performance may be linked to an individual's diligence and commitment to their work and the time spent 'on task' as discussed by Gibbs [195]. One avenue for future work would be to explore the 'diligence' of students and any potential gender differences that may exist in this area. Exploring the way in which students study or prepare for continual assessments and examinations may shed light on the observed gender disparity witnessed in physics, chemistry and biology results. In light of the observed gender performance gap, first year physics students in the 2012-13 academic year at Edinburgh were presented with a survey on study processes [196]. Results suggested that our students had a predominately 'strategic' or 'achieving' approach to their learning, which has been described by the literature as learning involving the use of "*any technique that achieves highest grades*" which results in a "*level of understanding patchy and variable*" [196]. This result was not surprising as it relates to a focus on studying with the aim of completing end-of-course examinations, but it does present a challenge when students are presented with university style teaching which focuses more on applying techniques to a range of physics problems set in a variety of contexts. A study skills intervention was piloted in the 2012-13 academic year, inviting students to take part in voluntary sessions in which they were given study skills materials that were not subject specific. Unfortunately the uptake for this pilot programme was very low, both when students in the lowest FCI performance quartiles were targeted and when it was subsequently opened up to the whole class. One suggestion for the future would be to incorporate this into class time or personal tutor meetings. By encouraging a move towards a deeper approach to learning, students may improve their overall understanding of physics concepts which could in turn enable them to transfer this knowledge to physics problems in a range of contexts. It is important for instructors to be aware that differences exist in the learning strategies of their students.

As mentioned in the previous section, results from the latest two years of CLASS data have shown no statistically significant decrease in level of expert-like thinking of first year students. Repeating this analysis for additional year groups can determine whether this result is a true effect or whether it is due to statistical fluctuations. If these results were to be replicated in future academic years, it may suggest that by adopting the format of the 'flipped classroom', in which students are presented with, and work through, the course material prior to the lectures, any decrease in the attitudes of students towards studying physics is eliminated and students maintain the same level of expert-like thinking after a semester of teaching. It could further confirm that introducing more interactive engagement methods into undergraduate teaching can have the effect of minimising the observed decline in students' attitudes to studying physics, compared to more traditional instruction methods. The study comparing US and UK academics attitudes to studying and learning physics has highlighted potential cultural differences in physics expert-like thinking, as well as noting a significant difference between male and female UK academics. It would be interesting to undertake further comparisons with other academics, particularly a larger North American academic population, to further investigate where these potential attitudinal differences exist.

8.3 Implications for instructors

Having discovered gender differences in the three aspects of undergraduate students' degree experiences, the next step is to determine what can be done to address such imbalance and reduce the gender participation and performance gap. Actions taken should be beneficial to both gender cohorts and improvements in the gender gap should not be to the detriment of male students.

- Increasing the number of female students who choose to pursue a degree in physics is imperative if we want to minimise the gender participation imbalance in tertiary education and academic careers. In order to encourage more students generally to continue studying physics and other STEM subjects, emphasis could be placed on better informing students about career prospects at an earlier age. Increasing the profile of women role models could also have the effect of making science more approachable to a wider female population. It is also important to maintain this at undergraduate level, a time when some students make the decision to transfer out of physics into other science disciplines.

- Instructors should be aware that quantitative results collected by diagnostic tests, although a good indication of the whole class performance, may mask underlying misconceptions held by both male and female students. As a result further discussion with students may be required to fully understand where such misconceptions lie.

- Misconceptions surrounding fundamental physics concepts of force and motion can result in students being unable to transfer their knowledge of physics concepts from one physics problem to another. Instructors should be aware of the wording of problems and explanations and how incorrect use of technical language amongst students can reinforce misconceptions. When using Peer Instruction in lectures, the lecturer could listen for appropriate use of technical language (for example correct distinction between force, acceleration and velocity) and ensure that students' discussions are focused on the intended physics concepts tested by the question.

- It is important for instructors to understand the efficacy of new teaching methods on improving students' conceptual understanding. Using a variety of interacting teaching methods can help to improve performance of both genders.

- Results have shown that students often try to rationalise their thinking by comparing the context of a problem with a 'real world' situation or with their personal experiences. Using examples and applications familiar to both females and males and of how physics principles are applied in everyday situations could improve students' understanding. In turn this could evoke awareness in students that studying physics can be valuable in various future careers.

- It is important to increase student engagement and for instructors to be aware of different learning environments which benefit different groups of students. Providing the opportunity for students to discuss problems with each other can be beneficial, as has been seen in results for in-lecture PI episodes. Results have also indicated that female students outperform male students on continual assessments, and that both gender cohorts expressed a desire for courses to be assessed using multiple methods, rather than those based solely on a final examination.

- Student confidence plays an integral role in both their perceptions of studying a subject as well as their overall performance. This study has shown that female students, compared to males, have a lower confidence level in their ability when answering correctly, but a slightly higher confidence in an incorrect answer. An individual's confidence can affect the level of engagement they have with their studies and can have a subsequent effect on their attitude towards studying a subject. It is therefore vital that instructors are aware of this issue.

8.4 Conclusion

The issues surrounding gender in physics education are complex and results suggest that there are multiple factors contributing to the differences between male and female participation, performance and attitudes in undergraduate physics. The under-representation of females in university physics cohorts is widely reported. The consistency of the proportion of females in university physics courses over the last decade suggests that it is very difficult to influence change once students have chosen the subjects that will dictate their choices in tertiary education. Exploration of course demographics has indicated that, although the percentage of female students enrolled in physics courses at the University of Edinburgh remained much lower than in other STEM subjects, there existed no gender differences in students' reasons for choosing their degree programme. This suggests that in order to improve the gender equality of both secondary school and undergraduate populations, the gender issue needs to be targeted earlier in students' primary and secondary education, a point at which students are often influenced by stereotypes and can subsequently lose interest in the sciences.

The inclusion of interactive engagement techniques in introductory physics courses has had an overall positive effect on students' performance and on reducing the gender gap in conceptual understanding tests. However, results from this study have uncovered strong evidence of gender performance differences across different assessment methods despite the introduction of such teaching methodologies. The introductory physics course taken by students in their first semester of the physics degree made use of interactive engagement methods during lectures and tutorial workshops but nevertheless indicated a persistent gender gap on a test of Newtonian mechanics in favour of males. In contrast, analysis of coursework results across several years of the physics degree programme showed a consistently higher performance by females compared to males, whilst end-of-course examinations showed less of a gender discrepancy. These results continue to raise the question of the extent to which gender performance is influenced by the structure of the assessment administered.

Identifying weaknesses in male and female students' understanding of key physics concepts is vital to pinpoint areas that need addressing during instruction. Examining interview comments and multiple choice response profiles to Newtonian mechanics questions allowed the question of whether these observed gender differences were a result of overall low conceptual understanding or if they related to specific misconceptions to be investigated. Results indicated that answers were strongly rooted in common misconceptions in students' understanding of key concepts and that in some cases this level of understanding differed by gender. By ascertaining where such conceptual difficulties exist and how students' attitudes change over time, instructors can more effectively support students during their undergraduate degree.

With regards to students' perceptions of their studies, results from attitudinal surveys hinted that the encouraging lack of decline in recent years in Edinburgh students' attitudes towards studying physics may be linked to changes in the lecture structure. Overall, male and female physics students were seen to share similar views on their degree experiences but observations once again suggested that further emphasis on promoting potential career paths should be undertaken in early years of education. This thesis has developed a greater understanding of the areas in which gender differences exist in physics students' degrees. Its results have highlighted areas for further investigation into factors that may influence the lessening of the gender gap and therefore improve students' overall university experience.

Bibliography

[1] T. Gill and J. F. Bell. What Factors Determine the Uptake of A-level Physics? *International Journal of Science Education*, 35(5):753–772, 2013.

[2] T. Mujtaba and M. J. Reiss. Inequality in Experiences of Physics Education: Secondary School Girls' and Boys' Perceptions of their Physics Education and Intentions to Continue with Physics After the Age of 16. *International Journal of Science Education*, 35(11):1824–1845, 2013.

[3] L. E. Kost, S. J. Pollock, and N. D. Finkelstein. Characterizing the gender gap in introductory physics. *Phys. Rev. ST Phys. Educ. Res.*, 5(1), 2009.

[4] Z. Hazari. Gender differences in introductory university physics performance: The influence of high school physics preparation and affect. In *American Astronomical Society Meeting Abstracts*, volume 38 of *Bulletin of the American Astronomical Society*, page 976, 2006.

[5] Y. Maeda and S. Y. Yoon. A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: Visualization of rotations (PSVT: R). *Educational Psychology Review*, 25(1):69–94, 2013.

[6] J. C. Blickenstaff. Women and science careers: leaky pipeline or gender filter? *Gender and Education (2005)*, 17(4):369–386, 2005.

[7] M. C. Linn and J. S. Hyde. Gender, Mathematics, and Science. *Educational Researcher*, 18(8):17–27, 1989.

[8] M. Wells, D. Hestenes, and G. Swackhamer. A modeling method for high school physics instruction. *Am. J. Phys*, 63(7):606–619, 1995.

[9] A. Colley, C. Comber, and D. J. Hargreaves. School subject preferences of pupils in single sex and co-educational secondary schools. *Educational Studies*, 20(3):379–85, 1994.

[10] Z. Hazari, G. Sonnert, P. M. Sadler, and M. Shanahan. Connecting High School Physics Experiences, Outcome Expectations, Physics Identity, and Physics Career Choice: A Gender Study. *Journal of Research in Science Teaching*, 47(8):978–1003, 2010.

[11] Government Policy: Engaging the public in science and engineering, https://www.gov.uk/government/policies/engaging-the-public-in-science-and-engineering–3, December 2012, Accessed April 2014.

[12] The National STEM Centre, http://www.nationalstemcentre.org.uk/, Accessed April 2014.

[13] Institute of Physics (IOP), http://www.iop.org/policy/statistics/education/file_43198-3, Accessed June 2011.

[14] Institute of Physics (IOP) Girls in Physics Action Research Programme Interim Report (2012), http://www.stimulatingphysics.org/pdfs/gip-interim-final-nov-12-low-res.pdf, Accessed March 2014.

[15] Joint Council for Qualifications (JCQ) Examination results summer 2010, http://www.jcq.org.uk/national_results/index.cfm, Accessed June 2011.

[16] Higher Entries and Awards by Gender, 2010, www.sqa.org.uk/sqa/files_ccc/nh10.xls, Accessed June 2011.

[17] Advanced Higher Entries and Awards by Gender, 2010, www.sqa.org.uk/sqa/files_ccc/ah10.xls, Accessed June 2011.

[18] E. Gillibrand, P. Robinson, R. Brawn, and A. Osborn. Girls' participation in physics in single sex classes in mixed schools in relation to confidence and achievement. *International Journal of Science Education*, 21(4):349–362, 1999.

[19] Institute of Physics (IOP) It's Different for Girls: The influence of schools (2012), http://www.iop.org/education/teacher/support/girls_physics/file_58196.pdf, Accessed March 2014.

[20] J. S. Eccles and J. E. Jacobs. Social forces shape math attitudes and performance. *Signs: Journal of Women in Culture and Society*, 11(2):367–380, 1986.

[21] Institute of Physics (IOP) Accepted Applicants to Degree Courses in UK Higher Education Institutions: Statistical Report (2012), http://www.iop.org/publications/iop/2012/file_56466.pdf, Accessed March 2014.

[22] Institute of Physics (IOP), http://www.iop.org/policy/statistics/education/file_43230, Accessed June 2011.

[23] Women in Science and Engineering (WISE), http://wisecampaign.org.uk, Accessed June 2011.

[24] Girls in physics, IOP, http://www.iop.org/education/teacher/support/girls_physics/page_41593.html, Accessed June 2011.

[25] E. Smith. Women into science and engineering? Gendered participation in higher education STEM subjects. *British Educational Research Journal*, 37(6):993–1014, 2011.

[26] A. Martens, M. Johns, J. Greenberg, and J. Schimel. Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, 42(2):236–243, 2006.

[27] S. J. Spencer, C. M. Steele, and D. M. Quinn. Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35(1):4–28, 1999.

[28] C. Good, J. Aronson, and J. A. Harder. Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, 29(1):17–28, 2008.

[29] J. R. Shapiro and A. M. Williams. The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66(3-4):175–183, 2012.

[30] M. Johns, T. Schmader, and A. Martens. Knowing Is Half the Battle, Teaching Stereotype Threat as a Means of Improving Women's Math Performance. *Psychological Science*, 16(3):175–179, 2005.

[31] J. E. McAdam. The persistent stereotype: children's images of scientists. *Physics Education*, 25(2):102–105, 1990.

[32] M. G. Jones, A. Howe, and M. J. Rua. Gender differences in students' experiences, interests amd attitudes toward science and scientists. *Science Education*, 84(2):180–192, 2000.

[33] D. W. Chambers. Sterotypic images of the scientist: The draw-a-scientist test. *Science Education*, 67(2):255–265, 1983.

[34] M. Mead and R. Metraux. Image of the scientist among high-school students. *Science*, 126(3270):384–390, 1957.

[35] J. Aronson and C. M. Steele. Stereotypes and the fragility of academic competence, motivation, and self-concept. *Handbook of competence and motivation*, pages 436–456, 2005.

[36] R. K. Thornton and D. R. Sokoloff. Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *Am. J. Phys*, 66(4):338–352, 1998.

[37] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito. Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation. *Science*, 330(6008):1234–1237, 2010.

[38] S. Lauer, J. Momsen, E. Offerdahl, M. Kryjevaskaia, W. Christensen, and L. Montplaiser. Stereotyped: Investigating gender in introductory science courses. *CBE Life Sciences Education*, 12(30), 2013.

[39] L. E. Kost-Smith, S. J. Pollock, N. D. Finkelstein, G. L. Cohen, T. A. Ito, and A. Miyake. Replicating a self-affirmation intervention to address gender differences: Successes and challenges. *AIP Conference Proceedings*, 1413(1):231–234, 2012.

[40] P. R. Clance and S. A. Imes. The Imposter Phenomenon in High Achieving Women: Dynamics and Therapeutic Intervention. *Psychotherapy Theory, Research and Practice*, 15(3):241–247, 1978.

[41] G. Gibson-Beverly and J. P. Schwartz. Attachment, entitlement, and the impostor phenomenon in female graduate students. *Journal of College Counseling*, 11(2):119–132, 2008.

[42] R. Ivie and K. N. Ray. Women in Physics and Astronomy, 2005. *Amer. Inst. Phys Publication Number R-430.02*, 2005.

[43] G. A. Buck, V. L. P. Clark, D. Leslie-Pelecky, Y. Lu, and P. Cerda-Lizarraga. Examining the cognitive processes used by adolescent girls and women scientists in identifying science role models: A feminist approach. *Science Education*, 92(4):688–707, 2008.

[44] R. Hake. Interactive-engagement vs traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys*, 66(1):64–74, 1998.

[45] C. H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *Am. J. Phys*, 69(9):970–977, 2001.

[46] D. Hestenes, M. Wells, and G. Swackhamer. Force Concept Inventory. *Phys. Teach.*, 30:141–158, 1992.

[47] J. Libarkin. Concept Inventories in Higher Education Science. In *BOSE Conf*, 2008.

[48] Conceptual Tests in the Physical Sciences, http://bit.ly/ConceptTests, Accessed April 2014.

[49] R. Chabay and B. Sherwood. Qualitative understanding and retention. *AAPT Announcer*, 27(96), 1997.

[50] E. Mazur. Peer instruction: Getting students to think in class. In Edward F. Redish and John S. Rigden, editors, *The Changing Role of Physics Departments in Modern Universities, Part Two: Sample Classes*, AIP Conference Proceedings, pages 981–988. American Institute of Physics, 1997.

[51] R. W. Preszler, A. Dawe, C. B. Shuster, and M. Shuster. Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE-Life Sciences Education*, 6(1):29–41, 2007.

[52] I Halloun and D. Hestenes. The initial knowledge state of college physics students. *Am. J. Phys*, 53(11):1043–1055, 1985.

[53] M. Prince. Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3):223–231, 2004.

[54] J. K. Knight and W. B. Wood. Teaching more by lecturing less. *Cell biology education*, 4(4):298–310, 2005.

[55] M. Lorenzo, C. H. Crouch, and E. Mazur. Reducing the gender gap in the physics classroom. *Am. J. Phys*, 74(2):118–122, 2006.

[56] S. Pollock, N. Finkelstein, and L. Kost. Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Phys. Rev. ST Phys. Educ. Res.*, 3(1):010107, 2007.

[57] A. Madsen, S. B. McKagan, and E. C. Sayre. Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics-Physics Education Research*, 9(2):020121, 2013.

[58] M. K. Smith, W. B. Wood, W. K. Adam, C. Wieman, J. K. Knight, N. Guild, and T. T. Su. Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science*, 323(5910):122–124, 2009.

[59] K. E. Perez, E. A. Strauss, N. Downey, A. Galbraith, R. Jeanne, and S. Cooper. Does displaying the class results affect student discussion during peer instruction? *CBE-Life Sciences Education*, 9(2):133–140, 2010.

[60] M. T. H. Chi, N. Leeuw, M. H. Chiu, and C. LaVancher. Eliciting self-explanations improves understanding. *Cognitive science*, 18(3):439–477, 1994.

[61] C. Turpen and N. D. Finkelstein. Not all interactive engagement is the same: Variations in physics professors' implementation of peer instruction. *Phys. Rev. ST Phys. Educ. Res.*, 5(2):020101, 2009.

[62] J. Docktor and K. Heller. Gender differences in both force concept inventory and introductory physics performance. *AIP Conference Proceedings*, 1064(1):15–18, 2008.

[63] S. P. Bates and R. K. Galloway. The inverted classroom in a large enrolment introductory physics course: a case study. *HEA STEM Conference Proceedings*, http://www.heacademy.ac.uk/events/detail/2012/academyevents/STEM_annual_conf 2012.

[64] S. P. Bates, R. Galloway, and K. McBride. Student-generated content: using peer-wise to enhance engagement and outcomes in introductory physics courses. volume 1413, pages 123–6. AIP Conference Proceedings, 2011.

[65] P. Laws, P. Rosborough, and F. Poodry. Women's responses to an activity-based introductory physics program. *Am. J. Phys*, 67(s1):s32–s37, 1999.

[66] J. Elwood. Equity issues in performance assessments: The contribution of teacher-assessed coursework to gender-related differences in examination performance. *Educational Research and Evaluation*, 5(4):321–344, 1999.

[67] M. Stewart. Gender issues in physics education. *Educational Research*, 40(3):283–293, 1998.

[68] R. Woodfield, S. Earl-Novell, and L. Solomon. Gender and mode of assessment at university: should we assume female students are better suited to coursework and males to unseen examinations? *Assessment & Evaluation in Higher Education*, 30(1):35–50, 2005.

[69] G. Stobart, J. Elwood, and M. Quinlan. Gender bias in examinations: how equal are the opportunities? *British Educational Research Journal*, 18(3):261–276, 1992.

[70] R. N. Steinberg and M. S. Sabella. Performance on multiple-choice diagnostics and complementary exam problems. *Phys. Teach.*, 35:150–155, 1997.

[71] N. Bolger and T. Kellaghan. Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, 27(2):165–174, 1990.

[72] J. Stewart, H. Griffin, and G Stewart. Context sensitivity in the force concept inventory. *Physical Review Special Topics-Physics Education Research*, 3(1):010102, 2007.

[73] P. Heller and M. Hollabaugh. Teaching problem solving through cooperative grouping. part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7):637–644, 1992.

[74] G. Taasoobshirazi and M. Carr. A review and critique of context-based physics instruction and assessment. *Educational Research Review*, 3(2):155–167, 2008.

[75] L. McCullough. Gender, Context, and Physics Assessment. *Journal of International Women's Studies*, 5(4), 2004.

[76] C. A. Ogilvie. Changes in students' problem-solving strategies in a course that includes context-rich multifaceted problems. *Phys. Rev. ST Phys. Educ. Res.*, 5(2):020102, 2009.

[77] K. D. Mattern and E. J. Shaw. A look beyond cognitive predictors of academic success: Understanding the relationship between academic self-beliefs and outcomes. *Journal of College Student Development*, 51(6):665–678, 2010.

[78] J. D. House. The relationship between self-belief, academic background and achievement of adolescent asian-american students. *Child Study Journal*, 27:95–110, 1997.

[79] S. Catsambis. Gender, race, ethnicity, and science education in the middle grades. *Journal of Research in Science Teaching*, 32(3):243–257, 1995.

[80] J. Osborne, S. Simon, and S. Collins. Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9):1049–1079, 2003.

[81] A. Baram-Tsabari and A. Yarden. Quantifying the gender gap in science interests. *International Journal of Science and Mathematics Education*, 9(3):523–550, 2011.

[82] E. F. Redish, R. N. Steinber, and Saul J. M. Student expectations in introductory physics. *Am. J. Phys*, 66(212-224), 1998.

[83] I. Halloun and D. Hestenes. Interpreting VASS dimensions and profiles for physics students. *Science Education*, 7:553–577, 1998.

[84] W. Adams, K. Perkins, N. S. Podolefsky, M. Dubson, N. Finkelstein, and C. Wieman. New instrument for measuring student beliefs about physics and learning physics: The colorado learning attitudes about science survey. *Phys. Rev. ST Phys. Educ. Res.*, 2(1):010101, 2006.

[85] H. Alhadlaq, F. Alshaya, S. Alabdulkareem, K. K. Perkins, W. K. Adam, and C. E. Wieman. Measuring students' beliefs about physics in saudi arabia. In *Physics Education Research Conference 2009*, volume 1179 of *PER Conference*, pages 69–72, 2009.

[86] K. Perkins, W. Adams, S. Pollock, N. Finkelstein, and C. Wieman. Correlating student beliefs with student learning using the colorado learning attitudes about science survey. In *Physics Education Research Conference 2004*, volume 790 of *PER Conference*, pages 61–64, Sacramento, California, August 4-5 2004.

[87] C. Baily and N. D. Finkelstein. Student perspectives in quantum physics. In *2008 Physics Education Research Conference*. American Institute of Physics, 2008.

[88] V. K. Otero and K. E. Gray. Attitudinal gains across multiple universities using the physics and everyday thinking curriculum. *Physical Review Special Topics-Physics Education Research*, 4(2):020104, 2008.

[89] E. Brewe, L. Kramer, and G. O'Brien. Modeling instruction: Positive attitudinal shifts in introductory physics measured with class. *Phys. Rev. ST Phys. Educ. Res.*, 5:013102, 2009.

[90] K. E. Gray, W. K. Adams, C. E. Wieman, and K. K. Perkins. Students know what physicists believe, but they don't agree: A study using the CLASS survey. *Phys. Rev. ST Phys. Educ. Res.*, 4:020106, 2008.

[91] W. G. Perry. JR. 1970. Forms of Intellectual and Ethical Development in the College Years: A Scheme. *New York: Holt, Rinehart and Winston*, 1970.

[92] The University of Edinburgh Factsheet 2012/2013: Student Figures.

[93] S. P. Bates, A. D. Bruce, and McKain D. Integrating e-learning and on-campus teaching: I. an overview, exploring the frontiers of e-learning: Research

proceedings of 12th association of learning technology conference alt-c, exploring the frontiers of e-learning, manchester. pages 130–9, 2006.

[94] S. P. Bates. Reshaping large-class undergraduate science courses: the weekly workshop. cal-laborate. 14:1–6, 2005.

[95] P. Denny, J. Hamer, A. Luxton-Reilly, and H. Purchase. Peerwise: students sharing their multiple choice questions. pages 51–58, 2008.

[96] G. M. Novak, E. T. Patterson, A. D. Gavrin, and W. Christian. *Just in Time Teaching*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[97] Submission for Juno Champion status, The School of Physics and Astronomy, The University of Edinburgh. November 2013.

[98] A. F. Heckler and E. C. Sayre. What happens between pre- and post-tests: Multiple measurements of student understanding during an introductory physics course. *Am. J. Phys*, 78(7):768–777, 2010.

[99] L. Ding, N. W. Reay, A. Lee, and L. Bao. Effects of testing conditions on conceptual survey results. *Phys. Rev. ST Phys. Educ. Res.*, 4(1):010112, 2008.

[100] S. Lin and C. Singh. Using ismorphic problems to learn introductory physics. *Phys. Rev. ST Phys. Educ. Res.*, 7(2):020104, 2011.

[101] L. Ding, R. Chabay, B. Sherwood, and R. Beichner. Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Phys. Rev. ST Phys. Educ. Res.*, 2:010105, 2006.

[102] D. Hestenes and M. Wells. A mechanics baseline test. *The Physics Teacher*, 30(3):159–166, 1992.

[103] I. Halloun and D. Hestenes. Common sense concepts about motion. *Am. J. Phys*, 53(11):1056–1065, 1985.

[104] D. Hestenes and I. Halloun. Interpreting the force concept inventory. *Phys. Teach.*, 33:502–6, 1995.

[105] I. Halloun and D. Hestenes. Revised and extended Force Concept Inventory, obtained via personal communication with authors.

[106] S. Ramlo. Validity and reliability of the force and motion conceptual evaluation. *Am. J. Phys*, 76(9):882–886, 2008.

[107] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx. Comparing the force and motion conceptual evaluation and the force concept inventory. *Phys. Rev. ST Phys. Educ. Res.*, 5(1):010105, 2009.

[108] Proceedings of the 2004 Physics Education Research Conference. *The Colorado Learning Attitudes about Science Survey*, number 790, 2005.

[109] J. Barbera, W. K. Adams, C. E. Wieman, and K. K. Perkins. Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *J. Chem. Educ*, 85(10):1435, 2008.

[110] K. Semar, J. K. Knight, G. Birol, and M. Smith. The Colorado Learning Attitudes about Science Survey (CLASS) for Use in Biology. *CBE Life Sci Educ*, 10:268–278, 2011.

[111] K. E. Gray, W. K. Adams, C. E. Wieman, and K. K. Perkins. Students know what physicists believe, but they don't agree: A study using the CLASS survey. *Phys. Rev. ST Phys. Educ. Res.*, 4:020106, 2008.

[112] A. Field. *Discovering statistics using SPSS*. Sage publications, 2009.

[113] A. Kay, J. Hardy, and S. Bates. A level playing field? Students' experiences of assessment in stem disciplines. *HEA STEM Conference Proceedings*, 2012.

[114] N. Lasry, E. Mazur, and J. Watkins. Peer instruction: From harvard to the two-year college. *Am. J. Phys.*, 76(11):1066–1069, 2008.

[115] Joint Council for Qualifications (JCQ) A Level Examination results summer 2013, http://www.jcq.org.uk/examination-results/a-levels/a-as-and-aea-results-summer-2013, Accessed May 2014.

[116] Joint Council for Qualifications (JCQ): A-Levels, http://www.jcq.org.uk/examination-results/a-levels, Accessed March 2014.

[117] K. Gray, S. Rebello, and D. Zollman. The effect of question order on responses to multiple-choice questions. In *Physics Education Research Conference*, 2002.

[118] S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch, and N. R. Walet. Gender differences in conceptual understanding of Newtonian mechanics: A UK cross-institution comparison. *European Journal of Physics*, 34(2):421–434, 2013.

[119] D. Sands and A. L. Marchant. Enhanced conceptual understanding in first year mechanics through modelling. *New Directions in Teaching of Physical Sciences*, 8:22–26, 2012.

[120] M. Birch and N. Walet. An integrated approach to encourage student centered learning: a first course in dynamics. *New Directions*, 4:21–26, 2008.

[121] HEA STEM Conference. *Using online assessment to provide instant feedback*, 2012.

[122] S. J. Pollock. Longitudinal study of student conceptual understanding in electricity and magnetism. *Phys. Rev. ST Phys. Educ. Res.*, 5:020110, Dec 2009.

[123] M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. W. Chabay, L. Ding, M. P. Haugan, M. J. Marr, B. A. Sherwood, and M. F. Schatz. Tale of two curricula: The performance of 2000 students in introductory electromagnetism. *Phys. Rev. ST Phys. Educ. Res.*, 5:020105, 2009.

[124] S. J. Pollock. Comparing student learning with multiple research-based conceptual surveys: CSEM and BEMA. volume 1064, pages 171–174. AIP, 2008.

[125] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein. Gender disparities in second-semester college physics: The incremental effects of a "smog of bias". *Phys. Rev. ST Phys. Educ. Res.*, 6(2), 2010.

[126] A. M. L. Cavallo, W. H. Potter, and M. Rozman. Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors. *School Science and Mathematics*, 104(6):288–300, 2004.

[127] K. A. Shaw. The development of a physics self-efficacy instrument for use in the introductory classroom. In *2003 Physics Education Research Conference*, volume 720, pages 137–140. AIP Publishing, 2004.

[128] Z. Hazari, P. M. Sadler, and R. H. Tai. Gender differences in the high school and affective experiences of introductory college physics students. *The Physics Teacher*, 46(7):423–427, 2008.

[129] R. M. Ryan and E. L. Deci. Intrinsic and extrinsic motivations: Classical definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.

[130] B. F. Jones and L. Idol. *Dimensions of Thinking and Cognitive Instruction.* Routledge, 2013.

[131] J. Clement. Students' preconceptions in introductory mechanics. *Am. J. Phys*, 50(1):66–71, 1982.

[132] J. P. Smith III, A. A. Disessa, and J. Roschelle. Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The journal of the learning sciences*, 3(2):115–163, 1994.

[133] P. B. Kohl and N. D. Finkelstein. Effects of representation on students solving physics problems: A fine-grained characterization. *Phys. Rev. ST Phys. Educ. Res.*, 2(1):010106, 2006.

[134] D. Meltzer. Relation between students problem-solving performance and representational format. *Am. J. Phys*, 75(3):463–478, 2005.

[135] N. S. Rebello and D. A Zollman. The effect of distracters on student performance on the force concept inventory. *American Journal of Physics*, 72(1):116–125, 2003.

[136] D. E. Brown. Students' concept of force: the importance of understanding newton's third law. *Phys. Educ*, 24(6):353, 1989.

[137] C. Terry and G. Jones. Alternative frameworks: Newton's third law and conceptual change. *European Journal of Science Education*, 8(3):291–298, 1986.

[138] A. Elby. Helping physics students learn how to learn. *American Journal of Physics*, 69(S1):S54–S6, 2001.

[139] N. Lasry, S. Rosenfield, H. Dedic, A. Dahan, and O. Reshef. The puzzling reliability of the force concept inventory. *Am. J. Phys*, 79(9), 2011.

[140] D. Temizkan. The effects of gender on different categories of students' miscocneptions about force and motion. *MSc thesis, Middle East Technical University*, Accessed Feb 2014.

[141] H. M. Breland, O. Danos, H. D. Kahn, M. Y. Kubota, and M. W. Bonner. Performance versus objective testing and gender: An exploratory study of an Advanced Placement history examination. *Journal of Educational Measurement*, 31(4):275–293, 1994.

[142] M. Beller and N. Gafni. Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles*, 42(1-2):1–21, 2000.

[143] M. E. Martinez. A Comparison of Multiple-Choice and Constructed Figural Response Items. *Journal of Educational Measurement*, 28(2):131–145, 1991.

[144] A. Savinainen and V. Jouni. The force concept inventory as a measure of students conceptual coherence. *International Journal of Science and Mathematics Education*, 6(4):719–740, 2008.

[145] D. Rosengrant, A. Van Heuvelen, and E. Etkinga. Do students use and understand free-body diagrams? *Phys. Rev. ST Phys. Educ. Res.*, 5(1):010108, 2009.

[146] M. Parlett and D. Hamilton. Evaluation as illumination: A new approach to the study of innovatory programs. occasional paper. *University of Edinburgh, Centre for Research in the Educational Sciences*, 1972.

[147] J. Potter. *Discourse Analysis and Constructionist Approaches: Theoretical Background. In John T.E. Richardson (Ed) Handbook of qualitative research methods for psychology and the social sciences.* BPS Books, Leicester, 1996.

[148] D. Silverman. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interactions. (2nd Edition).* Sage, London, 2001.

[149] S. Kvale, M.V. Angrosino, R.S. Barbour, Gibbs G., Flick U., and Rapley T. *Designing Qualitative Research.* Sage, London, 2007.

[150] Livescribe, http://www.livescribe.com/en-us/, Accessed March 2014.

[151] K. A. Ericsson and H. A Simon. *Protocol Analysis. Verbal reports as data.* Cambridge, MA: MIT Press, 1985.

[152] T. J. Cottle. *Black testimony: the voices of Britain's West Indians.* Wildwood House, 1978.

[153] M. W. Van Someren, Y. F. Barnard, and J. A. C Sandberg. *The think aloud method: A practical guide to modelling cognitive processes.* Academic Press, London, 1994.

[154] J. Nielsen, T. Clemmensen, and C. Yssing. Getting Access to What Goes on in People's Heads?: Reflections on the Think-aloud Technique. In *Proceedings of the Second Nordic Conference on Human-computer Interaction*, NordiCHI, pages 101–110, 2002.

[155] E. G. Mishler. Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review*, 60(4):415–442, 1990.

[156] M. Hammersley. *The relationship between qualitative and quantitative research: paradigm loyalty versus methodological eclecticism. In J.T.E. Richardson (ed) Handbook of Qualitative Research Methods for Psychology and the Social Sciences.* BPS Books, 1996.

[157] B. Glaser and A Strauss. *The discovery of grounded theory: Strategies for qualitative research.* Aldine, Chicago, 1967.

[158] K. Charmaz. *Constructing grounded theory: A practical guide through qualitative research.* London: Sage, 2006.

[159] U. Flick. *An introduction to qualitative research.* Sage, 2009.

[160] M. McCloskey. Intutitive physics. *Scientific American*, 248(4):122–130, 1983.

[161] X. Liu and D. MacIsaac. An investigation of factors affecting the degree of naìve impetus theory application. *Journal of Science Education and Technology*, 14(1):101–116, 2005.

[162] M. Chi, P. J. Feltovich, and R. Glaser. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2):121–152, 1981.

[163] D. Huffman and P. Heller. What does the force concept inventory actually measure? *The Physics Teacher*, 33(3):138–143, 1995.

[164] D. Laurillard. Styles and approaches in problem-solving. *The experience of learning*, 2, 1997.

[165] F. Reif and S. Allen. Cognition for interpreting scientific concepts: A study of acceleration. *Cognition and Instruction*, 9(1):1–44, 1992.

[166] R. M. Felder, G. N. Felder, M. Mauney, C. E. Hamrin, and E. J. Dietz. A longitudinal study of engineering student performance and retention. III. Gender differences in student performance and attitudes. *Journal of Engineering Education*, 84(2):151–163, 1995.

[167] WISE Women in Science, Technology, Engineering and Mathematics: from Classroom to Boardroom UK Statistics 2012, http://www.wisecampaign.org.uk/files/useruploads/files/wise_stats_document_final.pdf, Accessed December 2013.

[168] M. Inzlicht and T. Ben-Zeev. A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11(5):265–271, 2000.

[169] R. A. Johnson and G. I. Schulman. Gender-role composition and role entrapment in decision-making groups. *Gender & Society*, 3(3):355–372, 1989.

[170] D. Bruff. *Teaching with classroom response systems: Creating active learning environments.* John Wiley& Sons, 2009.

[171] J. D. Singer and J. B. Willett. Methodological issues in the design of longitudinal research: Principles and recommendations for a quantitative study of teachers' careers. *Educational Evaluation and Policy Analysis*, 18(4):265–283, 1996.

[172] R. Donnelly, C. MacPhee, and S. Bates. The Performance gender gap in undergraduate physics. *HEA STEM Learning and Teaching Conference*, http://journals.heacademy.ac.uk/doi/abs/10.11120/stem.hea.2012.069, 2012.

[173] R. Galloway, M. Wallace, and A. Wood. Closing the loop: enhancing the effectiveness of peer instruction questions by observing student discussions. HEA STEM Conference, 2014.

[174] M. C. James and S. Willoughby. Listening to student conversations during clicker questions: what you have not heard might surprise you! *American Journal of Physics*, 79(1):123–132, 2010.

[175] A. K. Wood, R. K. Galloway, J. Hardy, and C. M. Sinclair. Analyzing Learning during Peer Instruction Dialogues - A Resource Activation Framework. *Phys. Rev. ST Phys. Educ. Res.*, 2014, Manuscript submitted for publication.

[176] W. Mau and R. Lynn. Gender differences in homework and test scores in mathematics, reading and science at tenth and twelfth grade. *Psychology, Evolution & Gender*, 2(2):119–125, 2000.

[177] K. A. Slaughter. *Mapping the transition- content and pedagogy from school through university.* PhD thesis, University of Edinburgh, May 2012.

[178] E. Gire, B. Jones, and E. Price. Characterizing the epistemological development of physics majors. *Phys. Rev. ST Phys. Educ. Res.*, 5(1):010103, 2009.

[179] K. K. Perkins and M. Gratny. Who Becomes a Physics Major? A Long-term Longitudinal Study Examining the Roles of Pre-college Beliefs about Physics and Learning Physics, Interest, and Academic Achievement. In *2010 Physics Education Research Conference*, volume 1289, pages 253–256. AIP Publishing, 2010.

[180] M. Milner-Bolotin, T. Antimirova, A. Noack, and A. Petrov. Attitudes about science and conceptual physics learninng in univeresity introductory physics courses. *Phys. Rev. ST Phys. Educ. Res.*, 7(2):020107, 2011.

[181] HESA - Higher Education Statistics Agency, http://www.hesa.ac.uk/stats-staff, Accessed March 2014.

[182] D. E. Meltzer. The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores. *Am. J. Phys*, 70(12):1259, 2002.

[183] Course Experience Questionnaire: Curtin University, 2010.

[184] National Student Survey (NSS), http://www.thestudentsurvey.com, Accessed May 2014.

[185] Closing Doors, IOP, http://www.iop.org/publications/iop/2013/file_62083.pdfl, Accessed March 2014.

[186] The ROSE Survey in Scotland - An Extention Survey of Younger Pupils, http://www.gla.ac.uk/media/media_51343_en.pdf, Accessed March 2014.

[187] A. A. DiSessa. Toward an epistemology of physics. *Cognition and instruction*, 10(2-3):105–225, 1993.

[188] R. Driver, H. Asoko, J. Leach, P. Scott, and E. Mortimer. Constructing scientific knowledge in the classroom. *Educational researcher*, 23(7):5–12, 1994.

[189] S. Vosniadou. Capturing and modeling the process of conceptual change. *Learning and instruction*, 4(1):45–69, 1994.

[190] W. G. Perry. *Forms of Intellectual and Ethical Development in the College Years: A Scheme.* Holt, Rinehart and Winston, 1970.

[191] K. S Taber. The mismatch between assumed prior knowledge and the learner's conceptions: a typology of learning impediments. *Educational Studies*, 27(2):159–171, 2001.

[192] E. Mazur. The make-believe world of real-world physics. Presented at Pearson Marketing Forum, 22nd February 2014.

[193] M. Martin. Emotional and cognitive effects of examination proximity in female and male students. *Oxford Review of Education*, 23(4):479–486, 1997.

[194] E. Mazur. *Peer Instruction: A User's Manual.* Prentice Hall, 1997.

[195] G. Gibbs and C. Simpson. Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, 1(1):3–31, 2004.

[196] R. A. Fox, I. C. McManus, and B. C. Winder. The shortened Study Process Questionnaire: An investigation of its structure and longitudinal stability using confirmatory factor analysis. *British Journal of Educational Psychology*, 71(4):511–530, 2001.

Publications

S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch and N. R. Walet. Gender differences in conceptual understanding of Newtonian mechanics: A UK cross-institution comparison. *European Journal of Physics*, 34(2):421-434, March 2013.

R. Donnelly, C. MacPhee and S. Bates. The Performance gender gap in undergraduate physics. *Proceedings of the HEA STEM Learning and Teaching Conference,* 2012.

Appendices

Appendix A

Force Concept Inventory (FCI)

This concept inventory was administered to first year undergraduate physics students at the University of Edinburgh between 2011-13. A full copy of the Force Concept Inventory can be found through contact with the survey authors [46].

In 2011-12 and 2012-13 a second version of the Force Concept Inventory, in which the question were reordered, was administered to first year physics students. The order of questions for the 'Random Order' Force Concept Inventory was:

Questions 13, 18, 5, 6, 21, 22, 23, 24, 12, 29, 30, 19, 25, 26, 27, 15, 16, 14, 3, 7, 4, 8, 9, 10, 11, 17, 1, 2, 28, 10

Appendix B

Revised and extended Force Concept Inventory (FCI_{ext})

A revised and extended version of the Force Concept Inventory was administered to first year undergraduate physics students at the University of Edinburgh between 2006-10, a copy of which can be found through contact with the survey authors [105].

Appendix C

Force and Motion Conceptual Evaluation (FMCE)

The Force and Motion Conceptual Evaluation [36] was administered to students in their second year of the physics undergraduate degree at the University of Edinburgh. A copy of the assessment can be found through contact with the survey authors [36].

Appendix D

Brief Electricity and Magnetism Assessment (BEMA)

The Brief Electricity and Magnetism Assessment was administered to students in their second year of the physics undergraduate degree at the University of Edinburgh. A copy of the assessment can be found through contact with the survey authors [49].

Appendix E

Colorado Learning Attitudes about Science Survey (CLASS)

The following section contains a copy of the Colorado Learning Attitudes about Science Survey (CLASS) [84]. This survey contains 42 statements categorised into eight categories as follows:

Personal Interest: Questions 3, 11, 14, 25, 28 and 30.

Real World Connection: Questions 28, 30, 35 and 37.

Problem Solving General: Questions 13, 15, 16, 25, 26, 34, 40 and 42

Problem Solving Confidence: Questions 15, 16, 34 and 40

Problem Solving Sophistication: Questions 5, 21, 22, 25, 34 and 40

Sense Making / Effort: Questions 11, 23, 24, 32, 36, 39 and 42

Conceptual Understanding: Questions 1, 5, 6, 13, 21 and 32

Applied Conceptual Understanding: Questions 1, 5, 6, 8, 21, 22 and 40

CLASS

(Colorado Learning Attitudes about Science Survey)

Name: ______________________________ **Student ID Number:** ____________________

Introduction
Here are a number of statements that may or may not describe your beliefs about learning physics. You are asked to rate each statement by selecting a number between 1 and 5 where the numbers mean the following:

1. Strongly Disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly Agree

Choose one of the above five choices that best expresses your feeling about the statement. If you don't understand a statement, leave it blank. If you have no strong opinion, choose 3.

Survey
1. A significant problem in learning physics is being able to memorize all the information I need to know.

Strongly Disagree	1	2	3	4	5	Strongly Agree

2. When I am solving a physics problem, I try to decide what would be a reasonable value for the answer.

Strongly Disagree	1	2	3	4	5	Strongly Agree

3. I think about the physics I experience in everyday life.

Strongly Disagree	1	2	3	4	5	Strongly Agree

4. It is useful for me to do lots and lots of problems when learning physics.

Strongly Disagree	1	2	3	4	5	Strongly Agree

5. After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic.

Strongly Disagree	1	2	3	4	5	Strongly Agree

6. Knowledge in physics consists of many disconnected topics.

Strongly Disagree	1	2	3	4	5	Strongly Agree

7. As physicists learn more, most physics ideas we use today are likely to be proven wrong.

Strongly Disagree	1	2	3	4	5	Strongly Agree

8. When I solve a physics problem, I locate an equation that uses the variables given in the problem and plug in the values.

Strongly Disagree	1	2	3	4	5	Strongly Agree

9. I find that reading the text in detail is a good way for me to learn physics.

Strongly Disagree	1	2	3	4	5	Strongly Agree

10. There is usually only one correct approach to solving a physics problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree

11. I am not satisfied until I understand why something works the way it does.

Strongly Disagree	1	2	3	4	5	Strongly Agree

12. I cannot learn physics if the teacher does not explain things well in class.

Strongly Disagree	1	2	3	4	5	Strongly Agree

13. I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations.

Strongly Disagree	1	2	3	4	5	Strongly Agree

14. I study physics to learn knowledge that will be useful in my life outside of school.

Strongly Disagree	1	2	3	4	5	Strongly Agree

15. If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works.

Strongly Disagree	1	2	3	4	5	Strongly Agree

16. Nearly everyone is capable of understanding physics if they work at it.

Strongly Disagree	1	2	3	4	5	Strongly Agree

17. Understanding physics basically means being able to recall something you've read or been shown.

Strongly Disagree	1	2	3	4	5	Strongly Agree

18. There could be two different correct values for the answer to a physics problem if I use two different approaches.

Strongly Disagree	1	2	3	4	5	Strongly Agree

19. To understand physics I discuss it with friends and other students.

Strongly Disagree	1	2	3	4	5	Strongly Agree

20. I do not spend more than five minutes stuck on a physics problem before giving up or seeking help from someone else.

Strongly Disagree	1	2	3	4	5	Strongly Agree

21. If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.

Strongly Disagree	1	2	3	4	5	Strongly Agree

22. If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations.

Strongly Disagree	1	2	3	4	5	Strongly Agree

23. In doing a physics problem, if my calculation gives a result very different from what I'd expect, I'd trust the calculation rather than going back through the problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree

24. In physics, it is important for me to make sense out of formulas before I can use them correctly.

Strongly Disagree	1	2	3	4	5	Strongly Agree

25. I enjoy solving physics problems.

Strongly Disagree	1	2	3	4	5	Strongly Agree

26. In physics, mathematical formulas express meaningful relationships among measurable quantities.

Strongly Disagree	1	2	3	4	5	Strongly Agree

27. It is important for the government to approve new scientific ideas before they can be widely accepted.

Strongly Disagree	1	2	3	4	5	Strongly Agree

28. Learning physics changes my ideas about how the world works.

Strongly Disagree	1	2	3	4	5	Strongly Agree

29. To learn physics, I only need to memorize solutions to sample problems.

Strongly Disagree	1	2	3	4	5	Strongly Agree

30. Reasoning skills used to understand physics can be helpful to me in my everyday life.

Strongly Disagree	1	2	3	4	5	Strongly Agree

31. We use this statement to discard the survey of people who are not reading the questions. Please select agree-option 4 (not strongly agree) for this question to preserve your answers.

Strongly Disagree	1	2	3	4	5	Strongly Agree

32. Spending a lot of time understanding where formulas come from is a waste of time.

Strongly Disagree	1	2	3	4	5	Strongly Agree

33. I find carefully analyzing only a few problems in detail is a good way for me to learn physics.

Strongly Disagree	1	2	3	4	5	Strongly Agree

34. I can usually figure out a way to solve physics problems.

Strongly Disagree	1	2	3	4	5	Strongly Agree

35. The subject of physics has little relation to what I experience in the real world.

Strongly Disagree	1	2	3	4	5	Strongly Agree

36. There are times I solve a physics problem more than one way to help my understanding.

Strongly Disagree	1	2	3	4	5	Strongly Agree

37. To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.

Strongly Disagree	1	2	3	4	5	Strongly Agree

38. It is possible to explain physics ideas without mathematical formulas.

Strongly Disagree	1	2	3	4	5	Strongly Agree

39. When I solve a physics problem, I explicitly think about which physics ideas apply to the problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree

40. If I get stuck on a physics problem, there is no chance I'll figure it out on my own.

Strongly Disagree	1	2	3	4	5	Strongly Agree

41. It is possible for physicists to carefully perform the same experiment and get two very different results that are both correct.

Strongly Disagree	1	2	3	4	5	Strongly Agree

42. When studying physics, I relate the important information to what I already know rather than just memorizing it the way it is presented.

Strongly Disagree	1	2	3	4	5	Strongly Agree

Appendix F

FCI Major and Gender Graphs

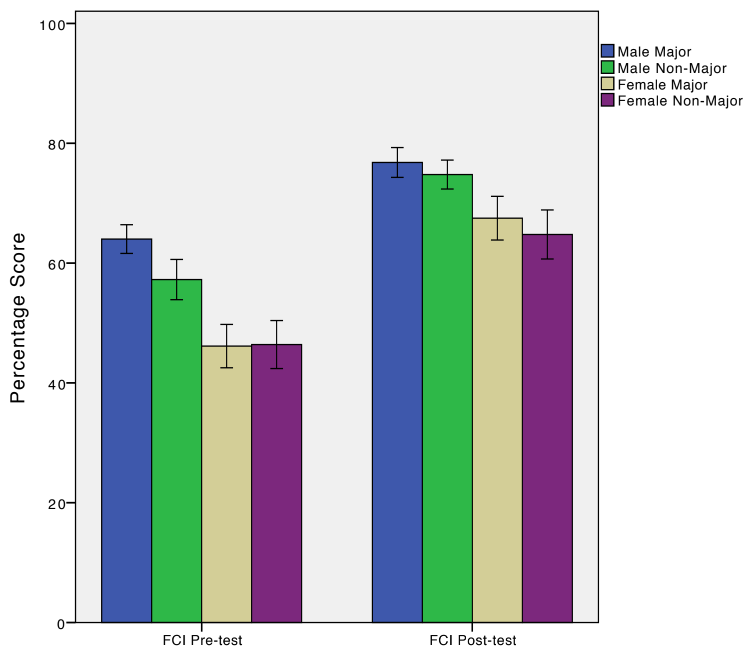


Figure F.1: Mean pre-test and post-test FCI scores for 2006-07 male and female majors and non-majors. Error bars represent the standard error on the mean.

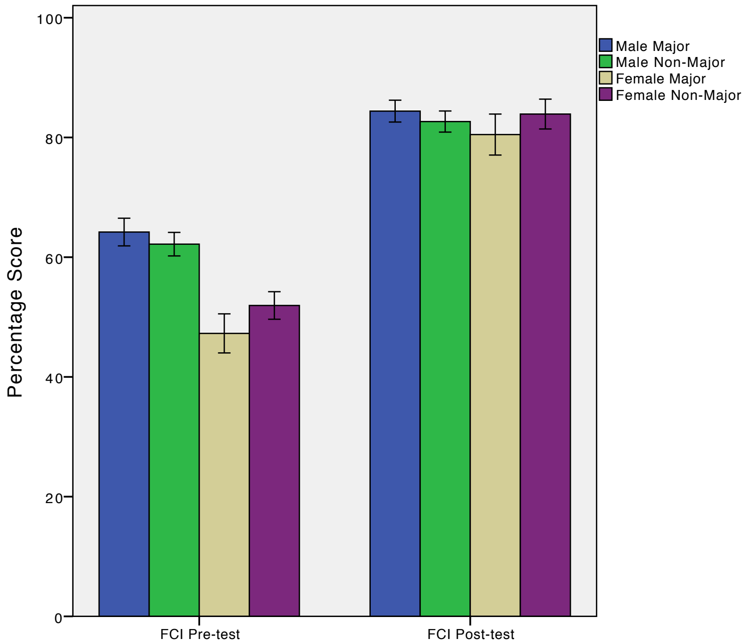


Figure F.2: Mean pre-test and post-test FCI scores for 2007-08 male and female majors and non-majors. Error bars represent the standard error on the mean.

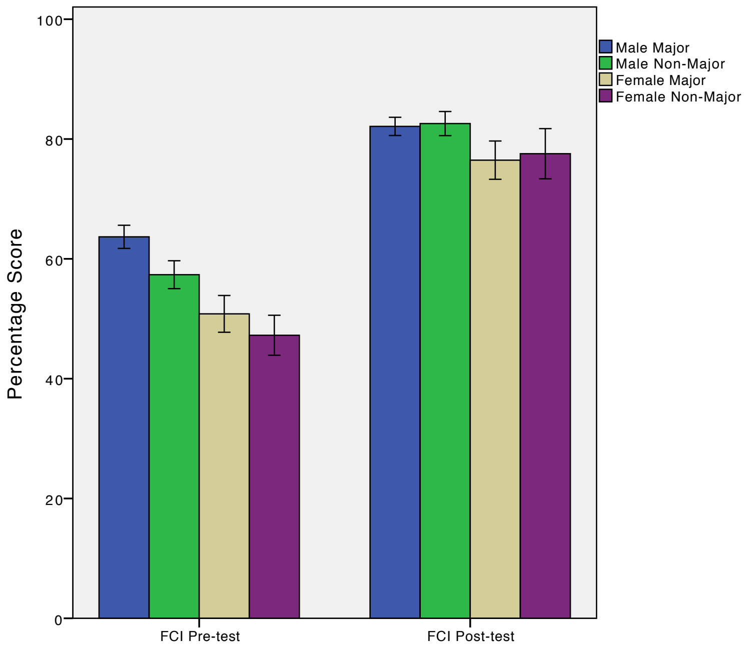


Figure F.3: Mean pre-test and post-test FCI scores for 2008-09 male and female majors and non-majors. Error bars represent the standard error on the mean.

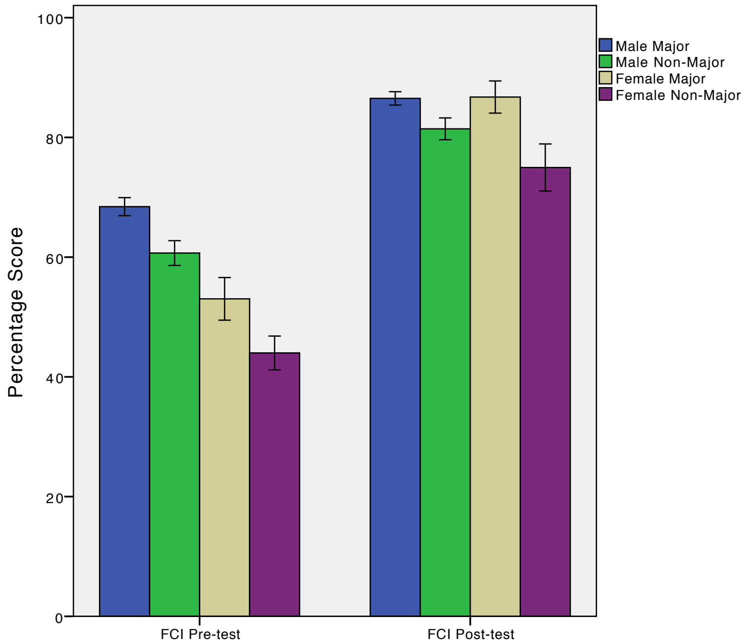


Figure F.4: Mean pre-test and post-test FCI scores for 2009-10 male and female majors and non-majors. Error bars represent the standard error on the mean.

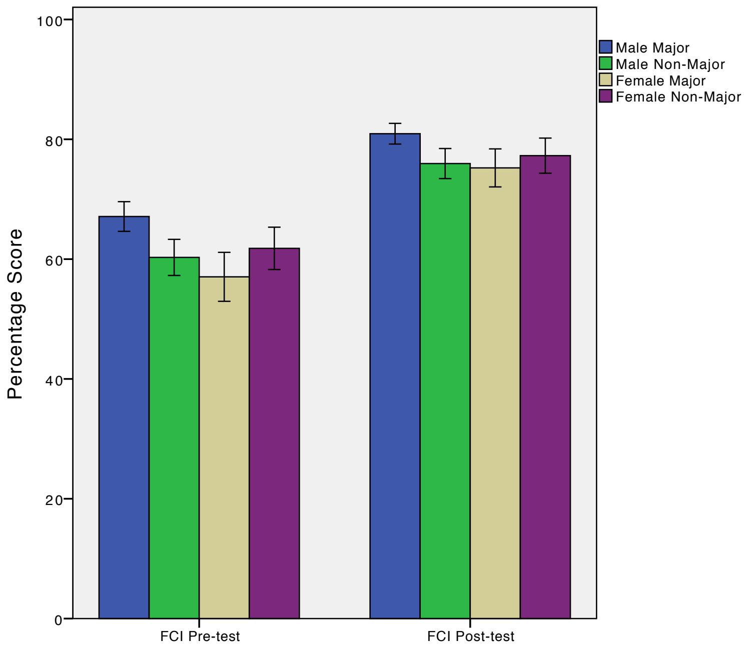


Figure F.5: Mean pre-test and post-test FCI scores for 2010-11 male and female majors and non-majors. Error bars represent the standard error on the mean.

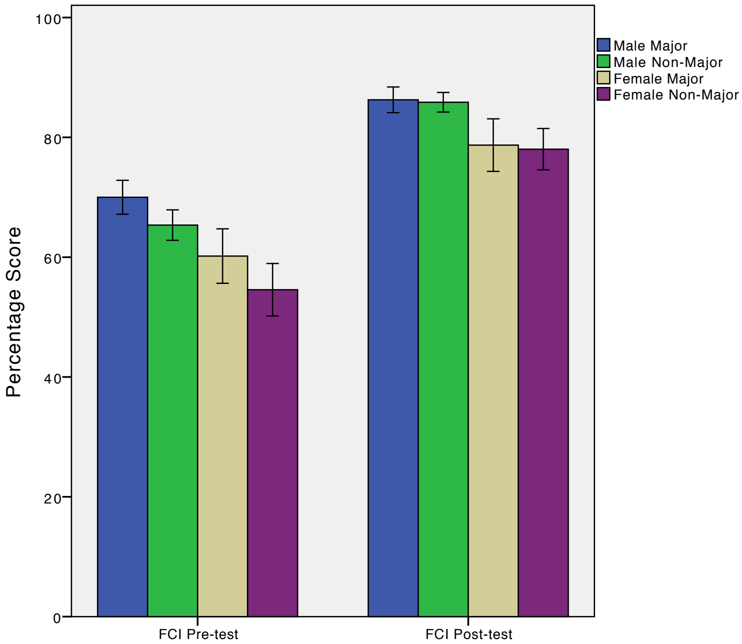


Figure F.6: Mean pre-test and post-test FCI scores for 2011-12 male and female majors and non-majors. Error bars represent the standard error on the mean.

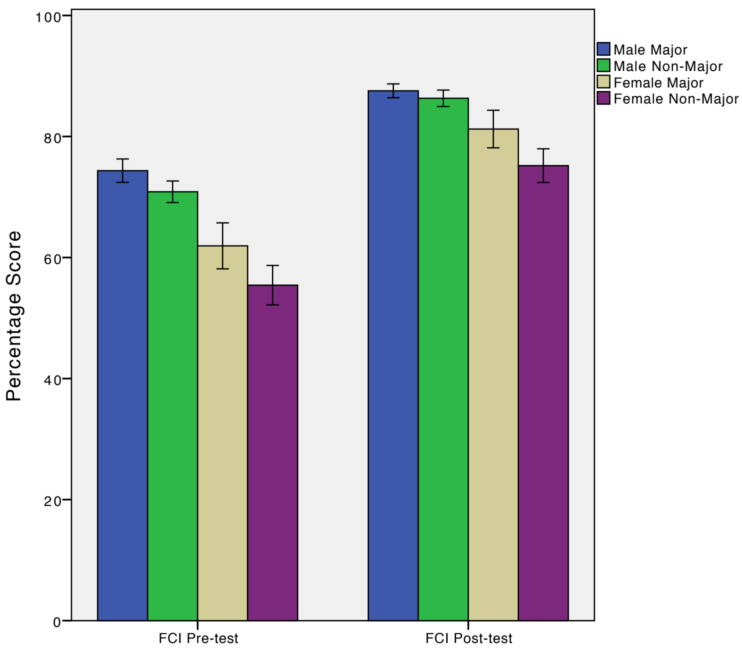


Figure F.7: Mean pre-test and post-test FCI scores for 2012-13 male and female majors and non-majors. Error bars represent the standard error on the mean.

Appendix G

FCI Quartile Distribution Graphs

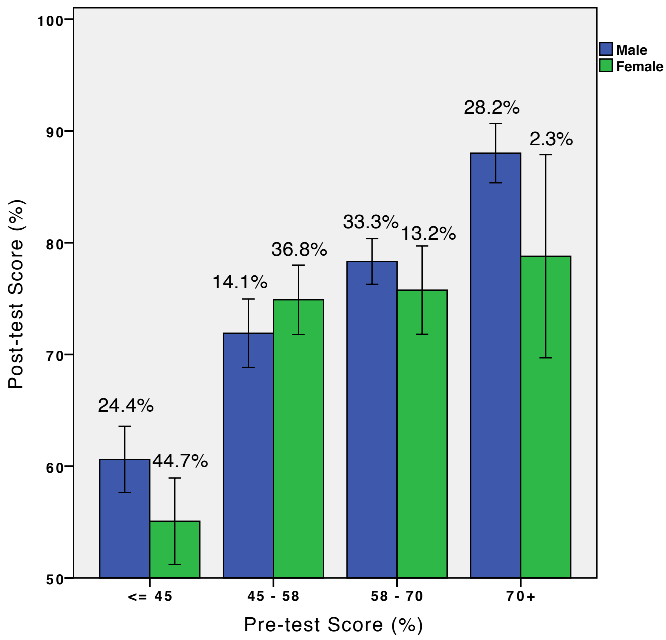


Figure G.1: FCI post-test scores as a function of male and female pre-test performance quartiles for 2006-07. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each cohort represented by each bar. N(males)= 78 and N(females)=38

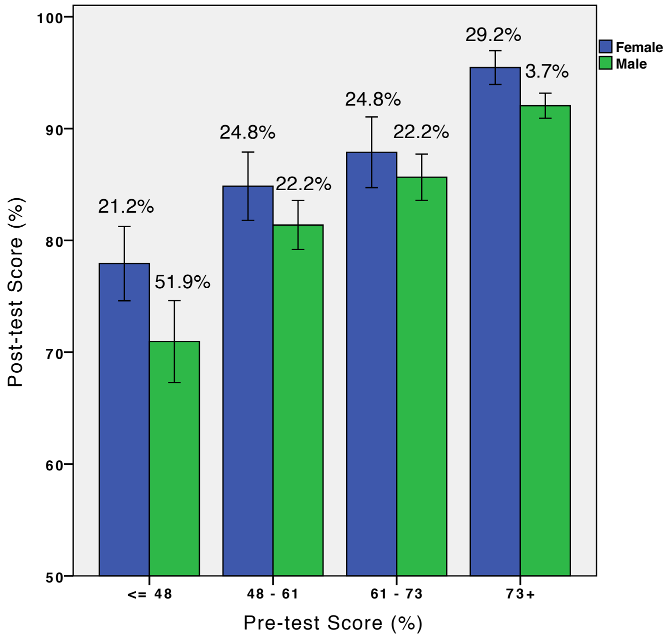


Figure G.2: FCI post-test scores as a function of male and female pre-test performance quartiles for 2007-08. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each cohort represented by each bar. N(males)= 137 and N(females)=54

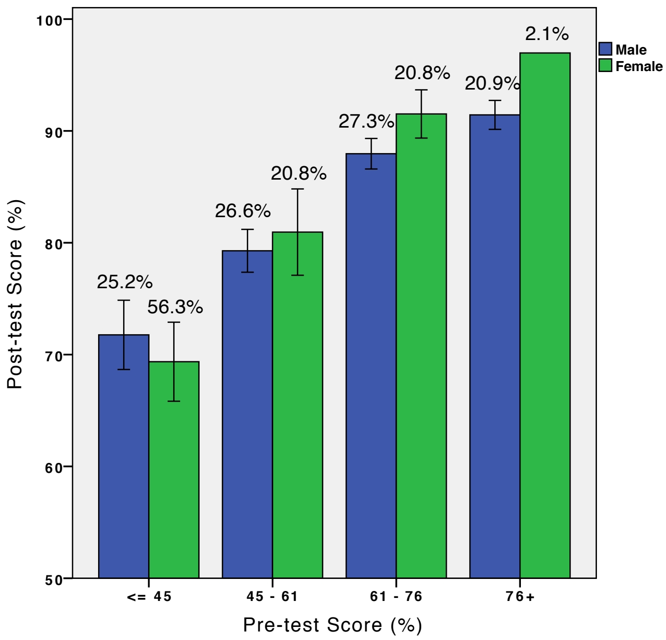


Figure G.3: FCI post-test scores as a function of male and female pre-test performance quartiles for 2008-09. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each cohort represented by each bar. N(males)= 139 and N(females)=48. Only one female student is in the top quartile.

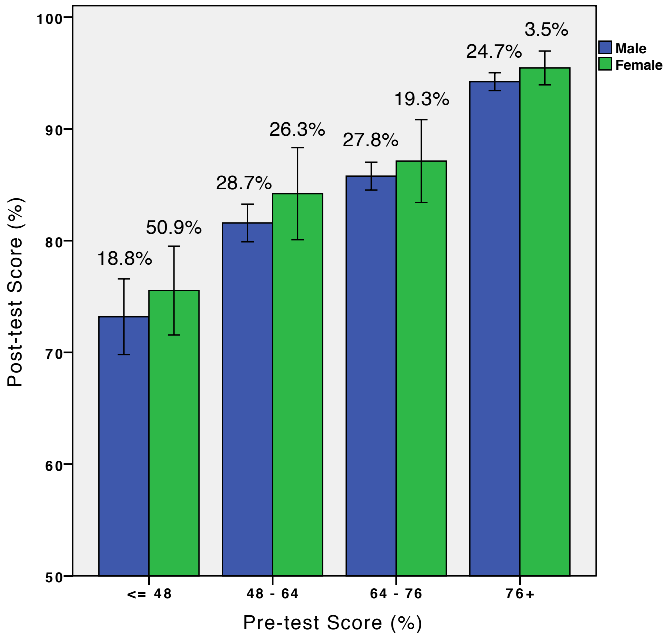


Figure G.4: FCI post-test scores as a function of male and female pre-test performance quartiles for 2009-10. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each cohort represented by each bar. N(males)= 223 and N(females)=57

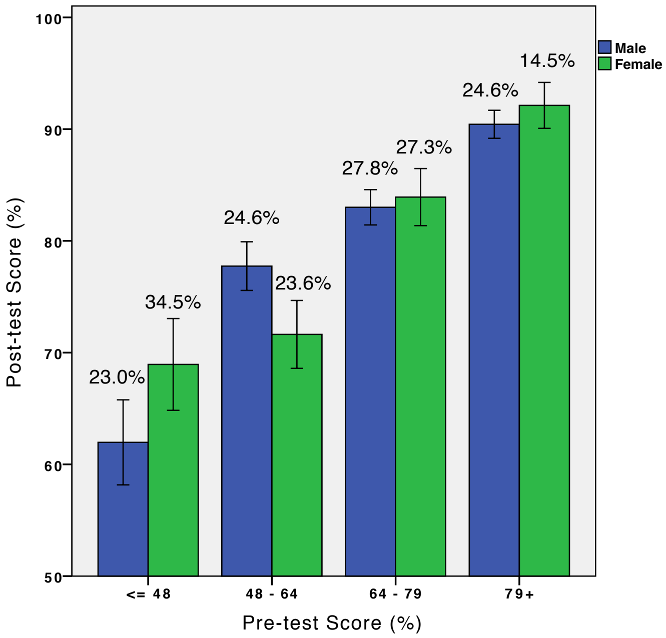


Figure G.5: FCI post-test scores as a function of male and female pre-test performance quartiles for 2010-11. Error bars represent the standard error on the mean. Percentage values above each bAar represent the percentage of students from each cohort represented by each bar. N(males)=126 and N(females)=55

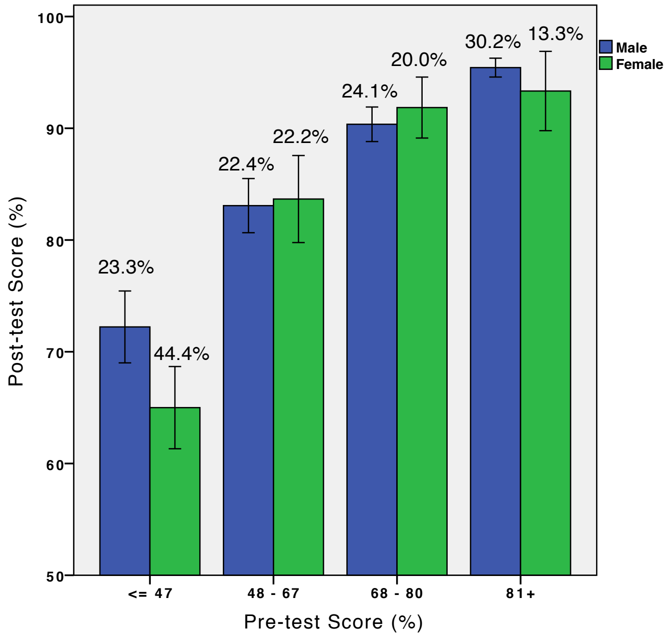


Figure G.6: FCI post-test scores as a function of male and female pre-test performance quartiles for 2011-12. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each cohort represented by each bar. N(males)= 116 and N(females)=45

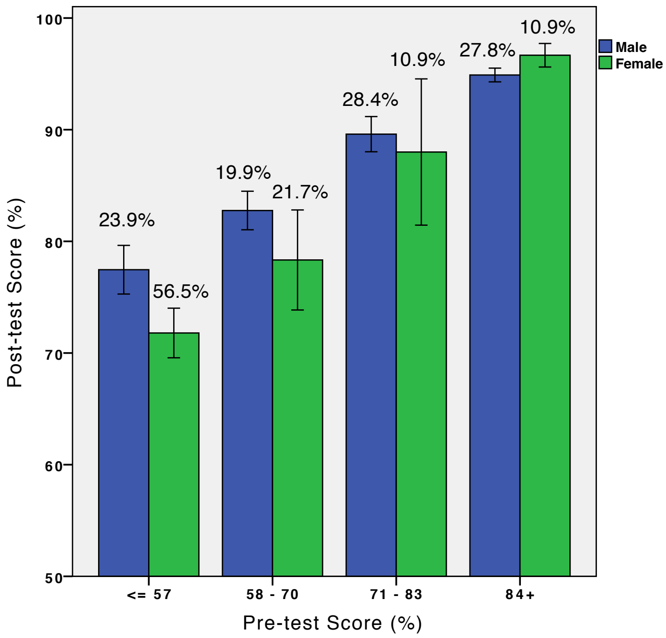


Figure G.7: FCI post-test scores as a function of male and female pre-test performance quartiles for 2012-13. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each cohort represented by each bar. N(males)= 176 and N(females)=46

Appendix H

Coursework and Examination Quartile Comparisons

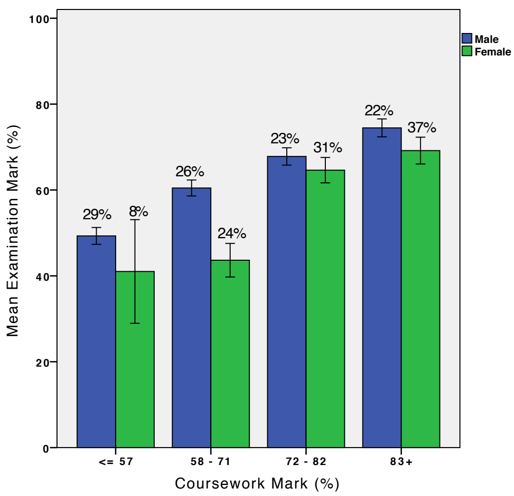


Figure H.1: End-of-course examination scores of 'Physics 1A' as a function of male and female coursework performance quartiles for the 2006-07 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=173 and N(females)=54.

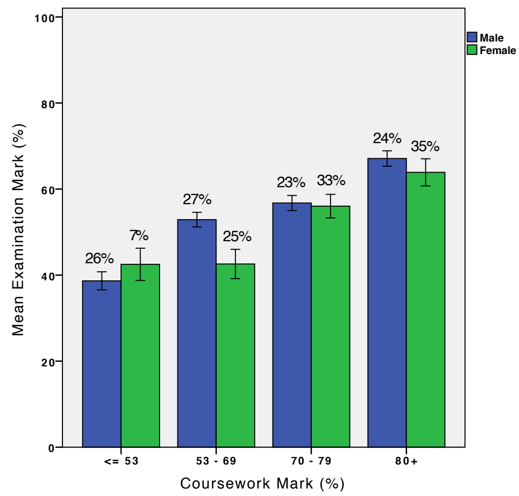


Figure H.2: End-of-course examination scores of 'Physics 1A' as a function of male and female coursework performance quartiles for the 2007-08 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=198 and N(females)=60.

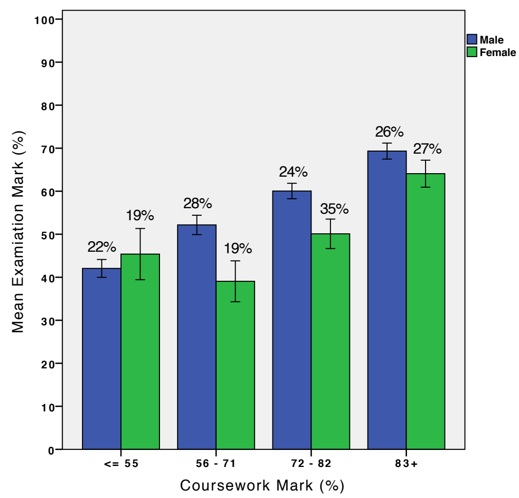


Figure H.3: End-of-course examination scores of 'Physics 1A' as a function of male and female coursework performance quartiles for the 2008-09 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=208 and N(females)=67.

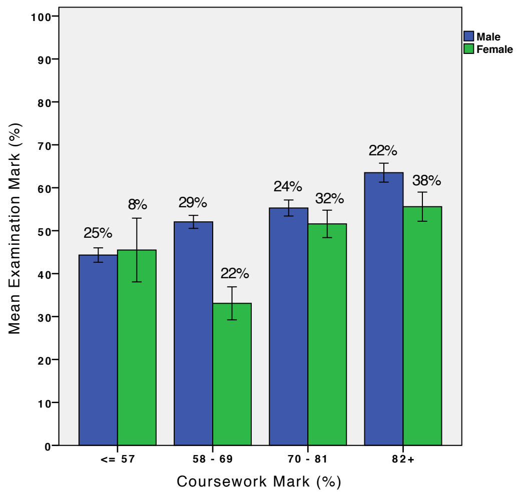


Figure H.4: End-of-course examination scores of ‘Physics 1A’ as a function of male and female coursework performance quartiles for the 2009-10 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=235 and N(females)=53.

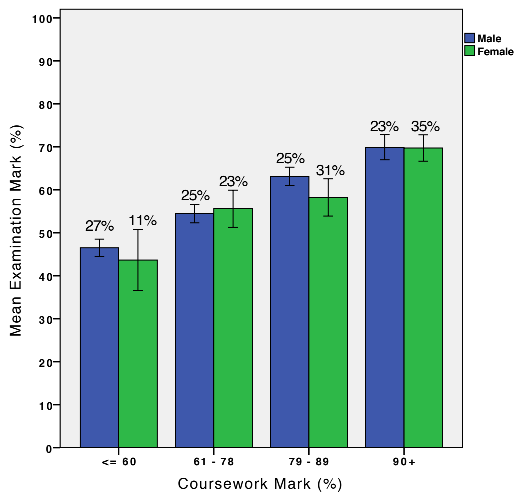


Figure H.5: End-of-course examination scores of ‘Physics 1A’ as a function of male and female coursework performance quartiles for the 2010-11 data set. Error bars represent the standard error on the mean. Percentage values above each bar represent the percentage of students from each gender cohort represented by each bar. N(males)=142 and N(females)=55.

Appendix I

CLASS Results

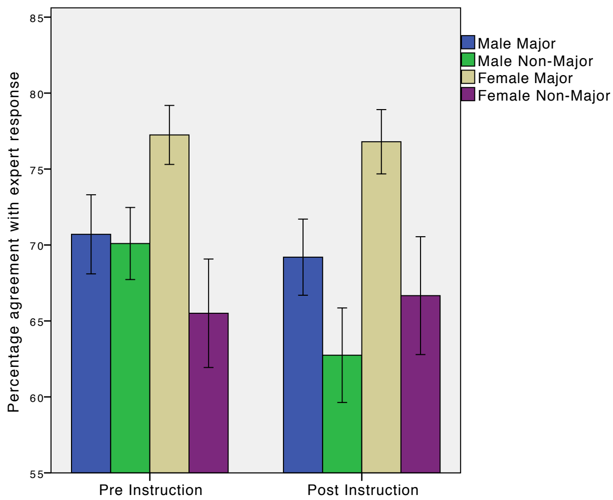


Figure I.1: Percentage of favourable expert-like thinking pre- and post-instruction as a function of gender and major for 2010-11 first year undergraduate students. N(male majors)=71 N(male non-majors)=34, N(female majors)=17 and N(female non-majors)=16.

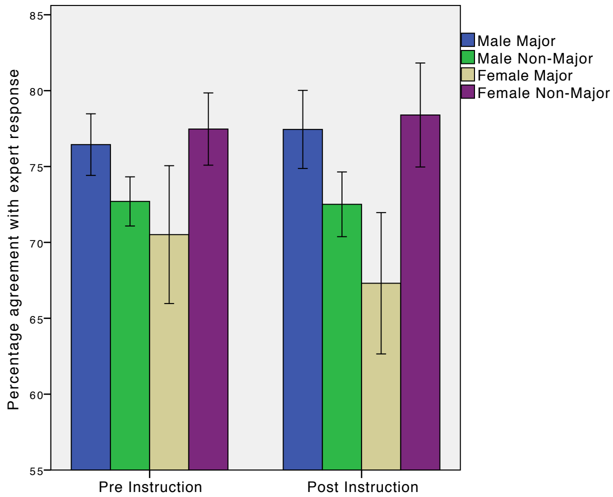


Figure I.2: Percentage of favourable expert-like thinking pre- and post-instruction as a function of gender and major for 2011-12 first year undergraduate students. N(male majors)=25 N(male non-majors)=29, N(female majors)=13 and N(female non-majors)=9.

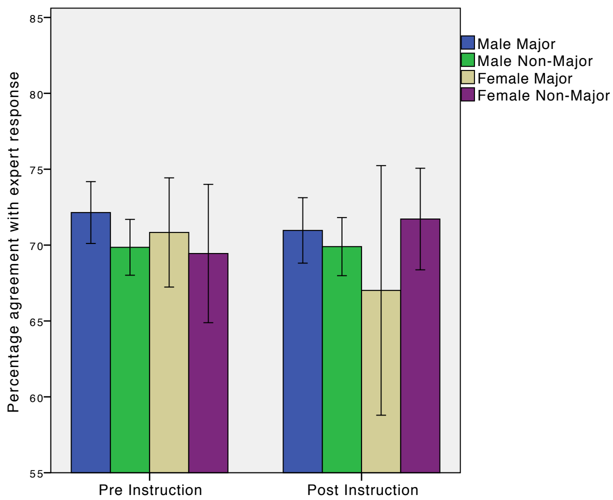


Figure I.3: Percentage of favourable expert-like thinking pre- and post-instruction as a function of gender and major for 2012-13 first year undergraduate students. N(male majors)=43 N(male non-majors)=61, N(female majors)=8 and N(female non-majors)=11.

Appendix J

Graduating Students Survey

The following three pages contain a copy of the survey presented to graduating students to study their attitude to their experiences of their undergraduate degree programme, the results of which are detailed in Chapter 7.

As part of an ongoing project within the Physics Education Research Group we are interested in gaining a greater understanding of students' perspectives on their experiences over the course of their degree studies.

It should take no more than 5-10 minutes to complete and your responses would be greatly appreciated. All responses will remain anonymous so please answer as openly and honestly as possible.

The following questions all ask you to rate your agreement / disagreement on a 5 point scale about your overall experiences across your entire programme of study

		Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
1	The staff put a lot of time into commenting on my work					
2	The teaching staff normally gave me helpful feedback on how I was going					
3	The programme helped me develop my ability to work as a team member					
4	It was always easy to know the standard of work expected					
5	The teaching staff on this programme motivated me to do my best work					
6	The programme provided me with a broad overview of my field of knowledge					
7	The programme sharpened my analytic skills					
8	My lecturers were extremely good at explaining things					
9	The teaching staff worked hard to make their subjects interesting					
10	The programme developed my confidence to investigate new idea					
11	The programme developed my problem-solving skills					
12	The staff made a real effort to understand difficulties I might be having with my work					

		Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
13	I usually had a clear idea of where I was going and what was expected of me in this course					
14	University stimulated my enthusiasm for further learning					
15	The programme improved my skills in written communication					
16	I learned to apply principles from this programme to new situations					
17	It was often hard to discover what was expected of me in this programme					
18	I consider what I learned valuable for my future					
19	As a result of my degree programme, I feel confident about tackling unfamiliar problems					
20	My degree programme helped me to develop the ability to plan my own work					
21	The staff made it clear from the start what they expected from students					
22	My university experience encouraged me to value perspectives other than my own					
23	Overall, I was satisfied with the quality of this programme					

24. What aspects of your degree programme were most in need of improvement?

____

____

____

____

25. What were the best aspects of your degree programme?

____

____

____

____

26. Any other comments you would like to make

____
____
____
____
____

What is your degree programme?

____

Are you enrolled on a BSc or Masters programme?

☐ BSc ☐ Masters

What are your future intentions after graduation?

Postgrad secured ☐ Employment secured ☐
Postgrad intended ☐ Employment Intended ☐
Undecided ☐ Other (please specify) ________________

What is your gender?

☐ Male ☐ Female

Thank you for completing this survey

If you have any questions about this survey feel free to contact

Robyn Donnelly
R.C.A.Donnelly@sms.ed.ac.uk
JCMB, Rm 1618
Kings Buildings
Edinburgh, EH9 3JZ
0131 650 6774